Zum Problem der Funktionsebenen bei projektiven Tests

untersucht am Beispiel der Wunschprobe.

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors der Philosophie
der Philosophischen Fakultät der Freien Universität Berlin

vorgelegt von

Waltraut Kerber-Ganse geb. Ganse

aus Stolpmünde/Pommern

1968

Druck:

ERNST-REUTER-GESELLSCHAFT
der Förderer und Freunde der Freien Universität Berlin e.V.

DISSERTATIONS-DRUCKSTELLE

1 Berlin 33, Garystr. 45

Zum Problem der Funktionsebenen bei projektiven Tests

untersucht am Beispiel der Wunschprobe.

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors der Philosophie
der Philosophischen Fakultät der Freien Universität Berlin

vorgelegt von

Waltraut Kerber-Ganse geb. Ganse

aus Stolpmünde/Pommern

UNIV.
LEIDEN
BIBL.

Dekan:	Prof. Dr. E. Loos
Referent:	Prof. Dr. H. Hörmann
Korreferent:	Prof. Dr. K. Holzkamp

Termin der mündlichen Prüfung: 11.März 1968

I N H A L T S V E R Z E I C H N I S

Seite

Einführung

1. Historischer Abriß zum Begriff der 'projektiven' Tests 7
2. Die Ansatzpunkte der Arbeit 21
3. Vorausschau auf den Gang der Darstellung 31

Erster Teil

Das Konzept der Funktionsebenen in seiner Relation zu anderen Forschungsansätzen in bezug auf projektive Tests

I. Das Problem der Funktionsebenen, soweit es in der Literatur explizit formuliert wird

1. Das Ebenen-Problem in seiner ersten Fassung : ROSENZWEIG 34
2. Das Ebenen-Konzept als Hilfskonstruktion zur Interpretation fehlender Validität von projektiven Tests 37
3. Das Ebenen-Konzept im Rahmen einer Persönlichkeitstheorie : LEARY 42
4. Vergleich des Ansatzes von LEARY mit dem der übrigen Autoren 48

II. Forschungsansätze zum Problem des Verhältnisses von Index und Indiziertem, soweit sie relevant sind für das Ebenen-Problem

1. Vorbemerkungen 54
2. Allgemeine Relationen zwischen Test-Index, Motiv und offenem Verhalten 55
3. Zur Genese der Testreaktion 61
   a) Zur Rolle von approach- und avoidance-Tendenzen 61
   b) Zur Rolle von perceptual sensitization und defense 74
4. Folgerungen aus den zitierten Untersuchungen 81

Zweiter Teil

Vorüberlegungen zu einem experimentellen Zugang zum Konzept der Funktionsebenen

III. Das Ebenen-Konzept als Rahmen einer construct-Validierung

1. Vorbemerkungen 92
2. Zum Problem fehlender Analysen der stimulus-relevance 93
3. Theoretische Vorklärungen zum Verhältnis von Ebenen-B e g r i f f und Test-O p e r a t i o n 99
4. Grenzen des Ebenen-Konzeptes 103

IV. Neufassung des Ebenen-Konzeptes

1. Das Ebenen-Konzept als Beitrag zur Test-Theorie 108
2. Das Ebenen-Konzept, verglichen mit anderen Persönlichkeitsmodellen 114

V. Konsequenzen für eine experimentelle Untersuchung

1. Vorbemerkungen 122
2. Allgemeine Charakterisierung der experimentellen Vorgehensweise 125
   a) Die Wunschprobe als experimenteller Ausgangspunkt 125
   b) Die Variablen des Experiments (Murray's need-system) 130
   c) Die Q_s-sort-Methode als Kriteriumsverfahren 134

VI. EXKURS : Die Q_s-sort-Methode in ihren theoretischen und experimentellen Implikationen

1. Vorbemerkungen 150
2. Zum Problem solipsistischer Kontinua 151
3. Zum Problem der intra-individuellen Vergleichbarkeit solipsistischer Kontinua 156
4. Zum Problem der inter-individuellen Vergleichbarkeit solipsistischer Kontinua 160
5. Zum Problem der forced method 165
6. Zum Problem der Repräsentativität der item-Stichprobe 172
7. Zum Problem der Variable 'social desirability' 179
8. Hypothesen zur Validität und Stabilität des hier verwendeten Q_s-sort-Verfahrens 184

Dritter Teil

D a s E x p e r i m e n t

VII. Die Durchführung des Experiments

1. Die Versuchsanordnung 190
2. Das Datenmaterial und seine statistische Verarbeitung 192
   a) Die Auswertung der Wunschprobe 192
   b) Die Auswertung des Q_s-sort-Verfahrens 202
   c) Die Auswertung der übrigen Verfahren 205

VIII. Die Ergebnisse des Experiments

1. Nachweis von Validität und Stabilität des Q_s-sort-Verfahrens 206
2. Aufweis verschiedener Funktionsebenen bei der WP 217
   a) Vorbemerkungen 217
   b) Zur Frage einer durchgängigen Gruppentendenz 225
   c) Zur Frage einer Extremgruppen-Differenzierung 227
   d) Zur Frage einer differentiellen Validität 242
   e) Vergleich der Ergebnisse der verschiedenen Auswertungsschritte 267
   f) Diskussion der Variablen nHarmavoidance 277
   g) Diskussion der Variablen nAggression 282

Schlußbetrachtung

1. Zusammenfassender Überblick über den Gang der Argumentation 292
2. Folgerungen aus den experimentellen Ergebnissen 297
3. Probleme und Konsequenzen einer differentiellen Validität 303

A N H A N G

1. Zur Wunschprobe 316

a) Das Auswertungsschema 316
b) Formale Hinweise für die Signierung 320
c) Tabelle : die Auswertung am Beispiel der need-scores einzelner Vpn 322

2. Zum Q_s-sort-Verfahren 323

a) Gesichtspunkte für die item-Konstruktion 323
b) Die items des Q_s-sort-Verfahrens 325

3. Instruktionen 333

a) Allgemeine Vorinstruktion 333
b) Instruktion zur Wunschprobe 333
c) Instruktion zum Q_s-sort-Verfahren 334
d) Instruktion zum Erinnerungsexperiment 337
e) Instruktion zum real-self(2) 337
f) Instruktion zum TAT 337

Verzeichnis der Tabellen 339

Verzeichnis der Abkürzungen 340

Literaturverzeichnis 341

E I N F Ü H R U N G

1. Historischer Abriß zum Begriff der 'projektiven' Tests

Das P h ä n o m e n der Projektion war schon sehr viel früher bekannt, als es die spätere fach-psychologische Diskussion vermuten läßt [1]: um so begründeter mag es daher erscheinen, wenn eine offenbar so allgemeingültige Tatsache zu einem zentralen Thema in einer Wissenschaft vom Menschen wird.

Der B e g r i f f der Projektion wurde erstmalig 1894 durch Freud psychologisch definiert [2]. Er besagt: Ich-bedrohliche Eigenschaften werden statt der eigenen Person einer fremden Person oder auch der Umwelt zugeschrieben; sie werden 'abgewehrt'. Dieser Begriff ging als 'klassische Projektion' in die Literatur ein. Eine weitere Fassung des Projektionsbegriffs führte Freud 1911 ein: 'Projektion wird nun sehr viel allgemeiner verstanden als 'Nach-Außen-Verlegen' innerer Vorgänge oder 'innerer Wahrnehmungen'[3].

Der Begriff der 'projective m e t h o d' schließlich wurde erstmalig bei Frank (1939) geprägt und umfaßt in der durch Frank gegebenen allgemeinen Form heute eine Vielzahl von Verfahren, die ihrer Entstehung nach bereits ein gutes halbes Jahrhundert umgreifen. So zählt der 'Rorschach', 1921 als 'Wahrnehmungstest' veröffentlicht (Rorschach 1921), heute zu den projektiven Tests, ähnlich wie es für die noch ältere Methode der Wort-Assoziation gilt.

Die ersten Arbeiten, die den Freud'schen 'klassischen' Begriff der Projektion als Abwehrmechanismus zur Grundlage hatten, wurden in den Jahren 1935/36 veröffentlicht, durch Morgan und Murray (1935), Sears (1936) und Cattell (1936). Murray ist dabei der erste, der den Vorgang der

1) So konstatiert schon Th.v.Kempen "What a man is inwardly, that will he be outwardly" (zit. n. Murstein/ Pryer 1959)
2) Freud 1961 (3.Aufl.) a, b
3) Freud 1961 (3.Aufl.) c, d, e

Projektion nutzbar zu machen versucht für ein d i a g n o s t i s c h e s Verfahren, das hinter die beobachtbare Oberfläche dringen soll, mit dem es also möglich ist "to search beyond, behind, beneath" (1937, S. 121).

Leitend ist bei allen sgn. projektiven Verfahren, die nun entwickelt werden (TAT,CAT,SET,PFT,WAT u.a. [1]), der Gedanke, daß relativ mehrdeutiges Reizmaterial zu Reaktionen führen müsse, welche besonders aufschlußreich sind für die individuelle Eigenart des Probanden. Da man die Genese dieser Reaktionen durch die Annahme von Projektionsvorgängen erklären zu können glaubte, erwartete man insbesondere den Ausdruck der verborgenen und konfliktbesetzten Bereiche der Person im Testmaterial. Die Interpretation dieses Testmaterials lasse sich, so glaubte man, durch diese Projektionshypothese theoretisch hinreichend fundieren. Durch eben diese Projektionshypothese aber war auch gerechtfertigt, daß das diagnostische Erfassen der 'projezierten' Inhalte sich nur auf die Intuition des Testinterpreten gründen konnte. Diesen Verfahren ist daher - so wie sie ursprünglich verstanden wurden - gemeinsam, daß sie nicht einzelne Fähigkeiten, welche entlang eines metrisch faßbaren Kontinuums variieren, m e s s e n , sondern daß sie in einem eher idiographischen Sinne die freie Entfaltung des Individuums hervorrufen.

Die projektiven Verfahren (und damit die Projektionshypothese selbst) blieben jedoch nicht unberührt von den zunehmend verfeinerten Methoden der experimentellen Psychologie. Hier wurde ein Instrumentarium angeboten, mit welchem mehr und mehr die Funktionsweise einzelner Tests kritisch analysiert und differenziert werden konnte.

So sieht auch später Murray - im Versuch einer Anpassung an den Stand der inzwischen gewonnenen wissenschaftlichen Erkenntnisse über diese Verfahren - sich gezwungen, die Annahme direkter Projektion aufzugeben. Er führt statt-

---

1) vgl. Murray 1938, Bellak,L. und Bellak,S.S. 1949, Rosenzweig et.al. 1947, Lossen, Schott 1952

dessen (1951) den sehr weit gefaßten Begriff der 'imaginative projection' ein : auch bewußte und akzeptierte, also nicht-verdrängte Inhalte, auch Vorstellungen und Erwartungen, selbst augenblickliche Empfindungen werden unter diesem Begriff subsumiert. Die Reaktion auf projektive Tests läßt sich nicht länger mit jenem durch Freud ursprünglich beschriebenen Projektionsprozeß gleichsetzen. Murray faßt aber seinen späteren Projektionsbegriff so weit und allgemein, daß er nicht länger eine 'Erklärung' des Zustandekommens (und damit der Bedeutung) einer Testreaktion liefern kann. Der Begriff dient nicht mehr einer präzisen Bezeichnung eines fest umrissenen Gegenstandes. Dann aber ist zu fragen, ob er nicht gänzlich aufzugeben ist.

Die Entwicklung, welche ausgehend von der ursprünglichen Projektionshypothese zu einer solchen unhaltbaren Erweiterung des Projektionsbegriffs führte, sei im folgenden kurz skizziert. Dieser Überblick soll deutlich machen, wie notwendig es ist, neue theoretische Ansätze zu finden. Einen solchen theoretischen Ansatz zu entwickeln, ist ja Aufgabe dieser Arbeit.

Murray schreibt zur Zeit der ersten Veröffentlichungen über den TAT : "The chief advantage of the indirect procedure based on projections und imaginal elaboration is that the subject believes that he is talking about the material 'out there'..... He does not suppose that he is speaking about himself." (1937, S.131) "The purpose of this procedure is to stimulate literary creativity und thereby evoke fantasies that reveal covert and unconscious complexes." (1938, S.530) Und ferner : "It was anticipated that in performence of this task a subject would necessarily be forced to project some of his own fantasies into the material and thus reveal his more prevailing tendencies." (1938, S.531)

Diese Bemerkungen Murray's zeigen, daß die Anlehnung an den klassischen Freud'schen Projektionsbegriff, nämlich den der abwehrmäßigen Projektion von Ich-bedrohlichen

Inhalten auf andere Personen, schon damals nicht sehr eng war. Vielmehr wurde bei diesen ersten Versuchen, die Funktionsweise eines projektiven Verfahrens theoretisch zu beschreiben, der allgemeinere, ebenfalls bei Freud formulierte Projektionsbegriff maßgebend : der des bloßen Nach-Außen-Verlegens 'gewisser Sinnesempfindungen' also die Projektion 'innerer Wahrnehmungen' o h n e die obige Bedingung von Abwehr oder Konflikt.

Schließlich wird dann der wenig später durch Frank (1939) geprägte allgemeine Begriff der 'projection method' noch weit unspezifischer gefaßt : "A projection method for study of personality involves the presentation of a ... stimulus-situation designed or chosen because it will mean to the subject, not what the experimenter has arbitrarily decided it should mean, but rather whatever it must mean to the personality who gives it, or imposes on it, his private, idiosyncratic meaning and organization." (S.403) Frank orientiert sogar seine Auffassung der projektiven Methoden überhaupt nicht am psychoanalytischen Projektionsbegriff. Er argumentiert vielmehr von Gestalt- und Feld-theoretischen Gesichtspunkten aus und vermeidet damit vollends eine Fundierung des Projektionsbegriffs selbst : er sagt nicht mehr, als d a ß projektive Tests in besonderer Weise den Erfordernissen einer solchen Persönlichkeitstheorie entgegenkämen. Die Frage jedoch, auf grund welcher Bedingungen ein projektiver Test die Persönlichkeit in dieser Weise widerspiegele, wird gar nicht berührt. Vielmehr wird in Lewin'scher Terminologie nur eine vage Behauptung aufgestellt, welche über das Verhältnis Test - Proband (oder Index - Indiziertes, siehe später) sehr viel weniger aussagt als jede noch so unkritische Anlehnung an das psychoanalytische Projektionskonzept : "A personality is the way an individual organizes and patterns life situations and effectively responds to them, 'structuralizes his life space', so that by projective methods we are evoking the very process of personality as it has developed to that moment." (S.403). Es ist dies eine wenig informative Verknüpfung von vermeintlicher Ursache und angeblicher Wirkung.

Auch später noch tauchen ähnlich unbegründete Erwartungen auf, so beispielsweise bei White (1944) : ausgehend davon, daß dem Probanden für die Wahrnehmung mehrdeutigen Materials keine konventionellen Muster zur Verfügung stehen, folgert er, daß die Inhalte dann nur aus der Person selbst stammen könnten : "..he (der Proband, K.-G.) is all but obliged to give expression to the most readily available forces within himself." (S.215)

Sind es nun die 'covert und unconscious complexes' (Murray 1938), sind es 'data on the process of organizing experinence' bzw. 'idiosyncratic patterns' (Frank, 1939) oder sind es 'the most available forces', (White, 1944), welche in den Reaktionen auf projektive Tests zum Ausdruck gelangen? Ist die Wirkungsweise bei allen projektiven Methoden die gleiche, oder fördern sie vielleicht verschiedene 'available forces' zu Tage? Welches aber wäre dann das Verhältnis der Aussagen verschiedener projektiver Verfahren zueinander? Welche Rolle spielt das in den Aussagen manifest Gegebene tatsächlich für die Person?

Es zeigt sich, daß die ersten Begründungsversuche der Wirkungsweise projektiver Verfahren widersprüchlich und insbesondere deshalb unzureichend sind, weil verschiedene theoretische Grundannahmen, die sich zur Beschreibung der 'projektiven Vorgänge' zwar anzubieten scheinen, unkritisch und ohne empirische Verifikation als tatsächliche 'Erklärung' mißverstanden werden. Allein die Annahme der klassischen Projektion liefert eine Hypothese über Bedingungen und Genese des Testverhaltens und damit die theoretische Grundlage für eine Interpretationsanweisung. Doch auch diese Annahme entbehrt jeglicher experimenteller Grundlagen, so daß zu untersuchen bleibt, wie weit sie einer eingehenden Prüfung standhält.

Es kamen allerdings diese Annahmen der sehr reichen, wie auch leicht zu erfassenden diagnostischen Relevanz der projektiven Verfahren den Bedürfnissen des klinischen Psychologen sehr weit entgegen. Gerade der Kliniker braucht Methoden, welche sowohl eine ganzheitliche Betrachtung der

Person, wie einen Zugang insbesondere zu den 'verborgenen', der direkten Befragung nicht zugänglichen Inhalten ermöglichen. So konnten Verfahren, welche gerade diese Erfordernisse zu erfüllen schienen, sich im Bereich der Praxis sehr schnell durchsetzen.

Vor den Anforderungen einer streng auf empirischer Prüfung insistierenden experimentellen Psychologie blieben sie - ihrer tatsächlich nicht zu leugnenden praktischen Bedeutung wegen - relativ lange bewahrt. Auch heute noch gibt es die Ansicht, daß die Eigenart projektiver Verfahren, vergleicht man sie mit psychometrischen Methoden, den Rückgriff auf Intuition erlaube. Diese Intuition ihrerseits läßt sich rechtfertigen durch die 'klinische' Erfahrung , - welche dem wissenschaftlichen Zugriff ebenso wenig zugänglich ist.

So sehr aber die Vielfalt des bei projektiven Methoden geäußerten Antworten-Materials auch dem Kliniker entgegenkommt und hier eben zu der unkritischen Annahme seiner unmittelbaren diagnostischen Relevanz geführt hat, so sehr ist gerade diese Mannigfaltigkeit das Dilemma des Experimentalpsychologen, welcher nur schwer einen Ansatz findet, den strengen Maßstab psychometrischer Reliabilitäts- und Validitätskriterien auf diese Tests anzuwenden. Zwar gibt es auch auf der klinisch-ganzheitlich orientierten Seite durchaus Validierungsuntersuchungen. Doch werden hier - dem klinischen Ansatz folgend - Test und Interpret als untrennbare Einheit validiert, was allen psychometrischen Forderungen nach 'Objektivität' oder 'Generalisierbarkeit' auf's schärfste zuwiderläuft. [1)]

Die projektiven Verfahren bieten so bis heute Anlaß zu sehr widerstreitenden Ansichten und stellen damit um so dringender die Aufgabe, einen angemessenen Zugang zu der ihnen eigenen Funktionsweise zu gewinnen. Es kommt hinzu, daß längst außer Zweifel steht, daß man von d e n pro-

1) für die verschiedenen Typen von Validierungsuntersuchungen siehe das Sammelreferat von Silverman 1959

jektiven Verfahren in irgend einheitlicher Weise nicht sprechen kann. Der Vorgang der Projektion selbst stellt nur einen geringen Teil der möglichen Reaktionsformen dar (vgl. Anderson 1951). Man kann projektive Tests weder danach vereinheitlichen, was sie indizieren, noch nach der Art des geforderten Reaktionstyps [1], noch nach der Art des dargebotenen stimulus-Materials. Am wenigsten aber werden sie durch die in ihrem gemeinsamen Oberbegriff implizierte Behauptung vereinheitlicht, daß diese Verfahren tatsächlich Projektionsvorgänge hervorriefen : gerade jene Annahme also ist von relativ geringer Bedeutung, welche bisher allein eine (scheinbar) begründete Interpretation projektiven Materials erlaubte [2].

Man ist also auch heute weit davon entfernt, eine einheitliche Theorie d e r projektiven Verfahren formulieren zu können. Vielmehr muß man annehmen, daß jeder Test auf grund der ihm eigenen Bedingungen auf der stimulus-Seite und auf der Reaktionsseite (s. Fußnote 1) eine je eigene Theorie verlangt : doch sind auch dafür bis heute die Kenntnisse bei weitem nicht ausreichend.

Erst in den 50iger Jahren hat eine intensive experimentelle Arbeit eingesetzt, sieht man von den wenigen frühen Untersuchungen ab [3], deren Konsequenzen speziell für die Frage des Funktionierens projektiver Tests erst später wieder aufgegriffen wurden. Hier, im Bereich der experimentellen Forschung, gibt es inzwischen eine Reihe fruchtbarer Ansätze, welche sich auf grund ihrer allgemein-psychologischen Orientierung von dem bisher so unergiebigen Projektionsbegriff haben frei machen können.

---

1) Die für einen bestimmten Test charakteristische (in der Instruktion geforderte) Reaktionsweise (z.B. eine Geschichte zu erfinden, einen Satz zu ergänzen oder Wörter zu assoziieren), sei im folgenden als 'Reaktionstyp' bezeichnet.

2) Auch läßt sich nicht sagen, daß alle sgn. projektiven Verfahren sich in vergleichbarer Weise von den psychometrischen Tests abhöben. Denn der PFT kommt einigen der psychometrischen Kriterien durchaus entgegen : standardisiertes Auswertungsschema, Validierung an inter-individuellen Normen.

3) Murray 1933; Sears 1936/37; Sanford 1936/37; Levine, Murphy 1942.

Es kommt nämlich hier erstmalig die Rolle des stimulus in den Blick : insbesondere auf dem Wege experimenteller stimulus-Variation konnte nun nachgewiesen werden, daß es nicht die 'Persönlichkeit' allein ist, welche die Reaktion determiniert. Der Vorgang der Projektion kann nun nicht mehr rein deskriptiv gefaßt und damit unbesehen vorausgesetzt werden : es wird vielmehr erkannt, daß der stimulus nicht als bloßer Auslöser für eine spezifisch individuelle Reaktion zu betrachten ist, sondern daß auch ihm eine entscheidende determinierende Rolle zukommt (siehe später).

Demgegenüber wird man von dem Versuch, den Projektionsbegriff dadurch zu halten, daß man das Projektionsgeschehen - in Anpassung an die inzwischen gewonnenen differenzierteren Einsichten - systematisch in verschiedene Formen aufgliedert, nicht wirklich neue und weiterführende Erkenntnisse erwarten können [1]. Es zeigen allerdings diese Versuche wenigstens so viel an, daß schon bei der - ja nicht durchgängig gerechtfertigten - Voraussetzung von Projektionsvorgängen dieser Art sehr verschiedene Spielarten für das Verhältnis der manifesten Testantwort zu dem durch sie indizierten psychischen Tatbestand gedacht werden können. Sie stellen damit durchaus Gesichtspunkte für die empirische Forschung bereit, nämlich verschiedene Hypothesen über denkbare Test-Reaktionen. Diese können - in einen anderen theoretischen Kontext übersetzt - als Leitgedanken für eine experimentelle Planung dienen - allerdings in einem relativ allgemeinen Sinne, da sie keine Hinweise auf operationale Korrelate enthalten.

Es sei dies am Beispiel der 4 Projektionsbegriffe von Murstein und Pryer(1959) erläutert (classical, attributive, autistic und rationalized projection). Es sollen dabei jedoch nicht jene Untersuchungen hier wiederholt werden, welche die Autoren selbst als Beleg der verschiedenen Projektionsformen heranziehen. Vielmehr soll hier nur kurz um-

1) So bei : Cattell 1951; Cattell, Wenig 1952; von Lennep 1957; Chase 1959; Murstein, Pryer 1959; Boesch 1960.

rissen werden, was erst an späterer Stelle ausführlicher behandelt wird: es soll aufgezeigt werden, wie wenig eine solche begriffliche Systematisierung, also ein Festhalten am Projektionsbegriff, der tatsächlichen Vielfalt experimentell aufgedeckter Reaktionen Rechnung zu tragen vermag. Der folgenden Diskussion experimenteller Ergebnisse kommt daher die (negative) Funktion zu, aufzuzeigen, daß der Projektionsansatz sich nicht in Hinsicht auf diese nachweislich 'andersartigen' Formen der Fantasie-Produktion exhaurieren [1] läßt. Diesem nur vorläufigen Abriß kommt nicht die positive Rolle zu, darzulegen, was heute als 'unbefragter Wissensbestand' gelten könnte.

Auf den Begriff der 'klassischen Projektion' lassen sich die Untersuchungen beziehen, welche den Ausdruck Ich-bedrohlicher Eigenschaften (z.B. hostility) in projektivem Material analysieren : und es zeigt sich hier, daß diese Eigenschaften mit größerer Wahrscheinlichkeit vermieden werden, als daß sie der Projektionshypothese gemäß in den 'Fantasie'-Reaktionen zum Ausdruck gelängen. Vor allem aber ergibt sich, daß eine Reihe von Faktoren für das Zustandekommen der einen oder anderen Reaktion maßgebend sind, welche in einem rein projektions-theoretisch orientierten Denken geradezu irrelevant erscheinen mögen : abgesehen von der schon entscheidenden Rolle der stimulus-Beschaffenheit sind es Dimensionen wie 'approach-avoidance', 'sensitization-defense', 'revealing- oder conceiling response-style', welche erst das Zustandekommen einer Reaktion bedingen und zu erklären vermögen (siehe Kap. II). Allerdings muß man Folgendes einräumen : wenn dem Vorgang der klassischen Projektion in den bisher durchgeführten Experimenten eine solche geringe Wahrscheinlichkeit zukommt, drückt sich darin nicht unbedingt eine 'objektive' Gewichtung aus; denn ein experimenteller Zugang zu den Bereichen, welche zugleich Ich-bedrohlich und unbewußt sind, ist nicht eben leicht zu

1) Dieser Begriff wurde von Dingler (1923) eingeführt und von Kaminski (1959,hier in einer gewissen Modifizierung) und von Holzkamp (1964) übernommen.

gewinnen. Zwar gibt es Arbeiten, die mit Hilfe soziometrischer Einschätzungsmethoden die Variable 'Selbsteinsicht-fehlende Selbsteinsicht' experimentell faßbar zu machen versuchen, doch zeigen gerade sie sehr widersprüchliche [1] Ergebnisse. Der Begriff der 'klassischen Projektion' stellt also durchaus noch seinerseits Aufgaben an die experimentelle Psychologie.

Auch für den Begriff der 'a t t r i b u t i v e n Projektion' steht eine Analyse der Bedingungen, welche zu einem solchen attributiven Nach-Außen-Verlegen innerer Gegebenheiten (wie Gefühl und Motivation) führen, heute noch aus. Theoretisch kann diese Lücke durch den Ansatz der social-perception-Forschung ausgefüllt werden. Doch gibt es von dieser Seite bisher keine Untersuchungen für den Ausdruck gerade von konsistenten oder generellen Persönlichkeits-Determinanten in den Testreaktionen, wie sie der Begriff der attributiven Projektion beschreibt. Die Theorie der social perception ist wohl - wie Hörmann (1954) und Graumann (1956) aufzeigen - der Eigenart des projektiven Testmaterials in augenfälliger Weise angemessen. [2] Doch hat sie nur zur Klärung eines ersten Vorverständnisses der Apperzeptionsverfahren beigetragen, nicht jedoch zu Hypothesen über deren Funktionsweise geführt. - Demgegenüber gibt es eine Fülle von Untersuchungen, welche durch eine einfache Punkt-zu-Punkt-Zuordnung von Testindex und Außenkriterium den direkten Motivausdruck in projektiven Verfahren nachweisen. [3] Zu verweisen ist hier aber auf ein Kon-

1) Nach Sears (1936,1944) projezieren Vpn mit fehlender Selbsteinsicht die betreffende Eigenschaft besonders stark auf andere Personen, ein Experiment, das u.a. von Murstein (1957a) überzeugend dahingehend kritisiert wird, daß es sich hier kaum um wirkliche Projektion handele. Abrams (1962) geht von der Hypothese aus, daß gerade die Gruppe mit Selbsteinsicht in eigene bedrohliche Eigenschaften mehr der betreffenden Inhalte bringt, nämlich sensitiver auf projektives Material reagiert, was wiederum für verschiedene Tests in verschiedenem Maße zu gelten scheint.

2) vgl. den Begriff der 'Hypothese' bei Bruner 1951, Postman 1951; Bruner und Postman 1951 u.a., und den Begriff des 'Schema' bei Vernon 1955

3) vgl. Kap. II

zept, nach welchem nicht für alle Motive ein attributiver Ausdruck erwartet werden kann, sondern insbesondere die unbefriedigten Motive 'hinausverlegt', nämlich in der Fantasie 'substitutiv' befriedigt werden [1]. Mit diesem 'alternative-channel model' (Broverman et.al.1960) erweist sich wiederum eine Differenzierung als notwendig, welche der Begriff der attributiven Projektion für sich genommen nicht leisten kann.

Auch bei der ' a u t i s t i s c h e n Projektion' sind noch nicht jene Bedingungen geklärt, welche dazu führen, daß die Umwelt, resp. die Test-stimuli so wahrgenommen werden, daß sie den eigenen gegenwärtigen Bedürfnissen entsprechen. Ein solcher Vorgang ist zwar experimentell aufgewiesen worden (Sanford 1936/37; Levine, Chein, Murphy 1942). Diesen Untersuchungen jedoch steht eine Reihe von Experimenten gegenüber, in welchen - bei vergleichbaren experimentellen Bedingungen - nicht eine autistische Projektion, sondern eine Vermeidung (inhibition) jeglicher bedürfnisrelevanter Aussagen überhaupt festzustellen war (Epstein 1962). In beiden Fällen wurde mit 'deprivation' (Entzug von Nahrung und oder Schlaf) gearbeitet : die Ergebnisse waren verschieden. Auch hier läßt sich also zeigen, daß andere theoretische Einsichten nötig sind, als sie der Begriff der Projektion vermittelt. (vgl. hierzu Kap. II)

Der Begriff der ' r a t i o n a l i z e d projection' bzw. der 'komplementären Projektion' (Allport 1939) findet bis heute am wenigsten eine Parallele in den bisherigen Untersuchungen. Es kann dies daran liegen, daß gerade diese Form von Projektion am stärksten handlungsrelevant ist : indem man etwa seine eigene offene Aggressivität dem Partner gegenüber dadurch fälschlich rechtfertigt, daß ein anderer einen ungebührlich frustriert habe. Wieweit in ähnlicher Weise auch die handlungs-irrelevanten Test-stimuli eine derartige Reaktion hervorrufen können, wäre eine interessante zukünftige Fragestellung. Alle bisherigen experimentellen Arbeiten,

1) vgl. Broverman et.al. 1960, Lazarus 1961; s.Kap. II.

welcher Provenienz auch immer, fragen stattdessen nach dem Grad der Relation zwischen Motiv und Motivausdruck : sie gehen von der Annahme aus, daß diagnostisch in erster Linie relevant sei, wieweit sich das Subjekt s e l b s t in einem Test widerspiegelt und nicht zugleich auch von der Frage, in welcher Art das Subjekt seine U m w e l t vermittels eines Tests charakterisiert und im Falle der komplementären Projektion gerade in ihren Intentionen dem Subjekt gegenüber falsch wiedergibt. [1)]

Kommt also dieser Abwehrmechanismus der Rationalisierung in der experimentellen Psychologie bisher nicht in den Blick, so ist es umgekehrt eine andere Form von Abwehr, welche erst in experimentellen Untersuchungen eingehend analysiert wurde, jedoch dem Konzept der Projektion direkt widerspricht : nämlich die Vermeidung (avoidance oder inhibition) von Motivausdruck überhaupt. Der Begriff der Projektion ist ja gerade dadurch definiert, daß etwas in bestimmter Weise 'nach außen verlegt wird', also zum Ausdruck gelangt. Die Vermeidung von Motiv-Ausdruck aber ist mit diesem Begriff theoretisch nicht faßbar, zumal 'Vermeidung' nicht etwa gleichgesetzt werden kann mit bloßer 'Auslassung' von Projektion. Das bloße Nicht-Auftreten von Projektion ist schon deshalb nicht von Bedeutung, weil es kein projektives Ver-

1) Diese Aufstellung von 4 Formen von Projektion, welche im übrigen der von Cattell (1951) entspricht (abgesehen davon, daß Cattell in Hinsicht auf Tests nicht von Projektion, sondern von 'misperception' spricht), wird von Chase (1960) in wenig überzeugender Weise kritisiert. Er lehnt den Begriff der rationalized projection aus rein formalistischen Gründen ab und kritisiert den der 'autistic projection' als viel zu allgemein auf jede Art von 'need-determined perception' generalisierbar, also als zu unspezifisch. Chase schlägt demgegenüber 2 Projektionsbegriffe vor : die defense- und die predictive projection (erstere etwa der klassischen, letztere etwa der attributiven Projektion entsprechend). Doch zeigt die obige Aufstellung, daß die 4 Projektionsbegriffe durchaus verschiedene Phänomene indizieren, so daß ihre Vermischung kaum weiterführend ist. Vielmehr wird der Begriff der 'predictive projection' bei Chase so allgemein gefaßt, daß auch die von Murstein (1959) gerade kritisierten Experimente, in denen Ähnlichkeiten zwischen der Selbstbeurteilung und der Beurteilung anderer fälschlich als Projektion aufgefaßt werden, ihm subsumiert werden können (vgl. auch Murstein 1961).

fahren gibt, das gezielt ganz bestimmte 'Projektionen' erwarten ließe, bei welchem also das F e h l e n erwarteter Projektionen überhaupt festgestellt werden könnte. Kann aber stattdessen nachgewiesen werden, daß ein Verfahren allgemein zu einer Thematisierung ganz bestimmter Inhalte führt, so ist deren Vermeidung u.U. von großer diagnostischer Relevanz. Das Phänomen der Vermeidung von Motivausdruck spielt daher heute im Rahmen der bisher aufgedeckten Funktionsmöglichkeiten projektiver Tests eine entscheidende Rolle. Nach einigen Autoren (siehe Kap. II) liegen die diagnostischen Möglichkeiten der projektiven Verfahren sogar ausschließlich in solchem Nachweis von Vermeidung. In diesem Gegenbegriff zur Projektion wird insbesondere deutlich, daß die Projektionshypothese eher die Vielfalt möglicher Funktionsweisen verschleiert, als daß sie als theoretisches Fundament für die projektiven Verfahren irgend aufschlußreich wäre.

Es trägt daher nicht eben zur Klärung bei, daß der Begriff 'projektiv' auch heute noch eingebürgert ist, obwohl er nur als ein historisches Relikt angesehen werden kann. So sehr auch Testreaktionen im nachhinein als 'Projektionen' charakterisiert werden können und so sehr es heute noch Aufgabe der experimentellen Psychologie ist, für die verschiedenen Formen von Projektion einen experimentellen Zugang zu erschließen, so ist doch der Sammelbegriff 'projektiv' nicht nur unzureichend, sondern irreführend. Von seiner (eingebürgerten) Konnotation her nämlich schwingt ungehindert die Vorstellung mit, daß hier auf einem quasi-mysteriösen Wege psychisch latente Inhalte sich dem diagnostischen Zugriff direkt präsentierten. Auf der anderen Seite jedoch, bei kritisch distanzierter Analyse dieser Verfahren, fällt man leicht viel zu [redacted]schnell in das andere Extrem der völligen Ablehnung dieser Tests, - dies vielleicht mit der Einschränkung, daß sie zu explorativen Zwecken von Nutzen sein mögen. Die weitreichenden Möglichkeiten aber einer experimentellen Analyse dieser Verfahren werden von beiden Seiten nicht gesehen.

Auch diese Arbeit kann den theoretischen Ansatz der Projektionshypothese nicht zu ihrem Ausgangspunkt machen : in dem ersten (nur vorläufigen) Überblick über den Stand der experimentellen Forschung ist deutlich geworden, daß die Analyse der Testreaktionen mit Hilfe der Forschungsmethoden der experimentellen Psychologie zu Ergebnissen führt, welche das theoretische Konzept der Projektion nicht mit zu umfassen vermag. Eine differenziertere Analyse des Testgeschehens ist daher notwendig. Zugleich sind theoretische Konzepte zu gewinnen, die nicht - wie beim Projektionsbegriff - nach dem Grad ihrer (vermeintlichen) Evidenz, sondern nach ihrem empirischen Erklärungswert zu beurteilen sind. Denn auf die kritische Rolle des Experiments als Kontrollinstanz für die Theorie kann insbesondere dort nicht verzichtet werden, wo die Theorie - wie es im Bereich der Diagnostik der Fall ist - Handlungsanweisungen zu begründen hat. [1] Der Begriff der 'Gültigkeit" eines Tests, verstanden als Begriff seiner 'empirisch begründeten Funktionsfähigkeit' (innerhalb einer fest umrissenen Aufgabenstellung), schließt notwendig derartige Handlungsanweisungen mit ein. Und eben diese verlangen nach einer Erfolgskontrolle, sollen sie Verbindlichkeit beanspruchen können. Vordringliche Aufgabe ist es daher, die Kenntnisse aus den verschiedenen Bereichen der experimentellen Psychologie weiter fruchtbar zu machen für diesen noch so problembehafteten Bereich innerhalb der psychologischen Diagnostik, damit er nicht wird, was Cattell (1951) warnend formulierte : "... 'the little boy who never grew up' among psychological techniques".

Das aber setzt voraus, daß man auf den Begriff 'projektiv' endgültig verzichtet, ihn weder durch einen anderen ersetzt [2] noch ihn möglichst eng, wenn auch präzise faßt [3],

---

1) im Gegensatz etwa zu axiomatisch-deduktiven Systemen

2) So geht z.B. Cattell (1951) vor mit seinem Begriff des dynaception test : es handelt sich hier um nicht mehr als einen deskriptiven Begriff, ohne d e n theoretischen Kontext, der zu empirischen Hypothesen führen könnte (vgl.S.).

3) Anderson (1951) z.B. behält den Begriff nur der 'klassischen Projektion vor und betont, daß dieser Vorgang nur einer unter sehr verschiedenartigen 'mental mechanisms' sei, welche durch projektive Tests hervorgerufen werden könnten, - ohne allerdings diese weiteren Mechanismen anzugeben.

noch daß man ihn dahingehend allgemein genug formuliert, daß er möglichst viele der heute nachgewiesenen Reaktionen auf (mehr oder weniger) mehrdeutiges Material umfassen möge, wie es 1951 von Murray vorgeschlagen wird : "To call something a projection, one must demonstrate, that it is similar to a 'relevant' constituent of the subject's personality. A relevant constituent is one that should be included in the final formulation of the personality" (S.XIV). Es impliziert allerdings diese Definition, die auf der einen Seite der Frank'schen Definition der projektiven Methoden an Unergiebigkeit nicht nachsteht, doch eine neuartige und entscheidende Einschränkung : es wird Projektion nicht mehr vorausgesetzt, sondern erst im nachhinein, auf der Grundlage der Testinterpretation, rückwirkend erschlossen. Das aber bedeutet bereits implicite, daß die Testinterpretation ohne Rekurs auf einen Projektionsbegriff vorgenommen werden kann (und muß). Der Projektionsbegriff, bei Murray ehedem zum Ausgangspunkt gemacht, hebt sich - in dieser Anpassung an den seither erreichten Stand der Forschung - nunmehr selbst auf.

Was aber kann an die Stelle der Projektionshypothese treten, welche umso mehr einer theoretischen Grundlage entbehrt, je mehr man versucht, sie den neueren experimentellen Erkenntnissen 'anzupassen' ? Dies zu beantworten ist bereits Aufgabe der Arbeit selbst und wird im Hauptteil ausführlich behandelt werden. Die Aufgabenstellung im engeren Sinne sei nun zunächst im folgenden präzisiert und ihr Ort innerhalb verschiedener wissenschaftlicher Richtungen genauer bestimmt.

2. Die Ansatzpunkte der Arbeit

Theoretische Kenntnisse und experimentelle Daten über die Funktionsweise 'projektiver' [1] Verfahren liegen heute in

1) Der Begriff 'projektiv' ist nunmehr so weit kritisiert und abgelehnt worden, daß im folgenden die solcher Einschränkung dienenden Anführungszeichen fortgelassen werden können; ein anderer Begriff wird deshalb nicht eingeführt, weil eine gemeinsame Bezeichnung aller dieser Verfahren sich gerade als fragwürdig erwiesen hat. Der Begriff wird hier also nur noch als 'Name' im eingebürgerten Sinne, doch ohne seine theoretischen Implikationen verwendet.

vielfältiger Form vor. Doch die Fülle verschiedenartiger Ansatzpunkte bringt es mit sich, daß bisher nur einzelne T e i l probleme analysiert wurden und eine Grundlage für ein einheitliches Theoriengebäude noch nicht gegeben ist. Es ist nicht allein die experimentelle Psychologie, welche sehr verschiedenartige und teils einander widersprechende Ergebnisse hervorgebracht hat. Die projektiven Verfahren sind Gegenstand auch verschiedener Disziplinen und Schulen der Psychologie : die Beiträge zu diesen Tests aus den verschiedenen Richtungen sind kaum übersehbar und weniger noch miteinander vereinbar. Dieser Umstand wurde von Hörmann (1964 a) sehr klar akzentuiert : "Tatsächlich ist hier (im Bereich der projektiven Verfahren, Verf.) der Ort, an welchem sich Psychoanalyse und Experimentalpsychologie, geisteswissenschaftliche und empirische Psychologie, ganzheitliche und einzelheitliche Vorgehensweise wirr und kaum durchschaubar überlagern." (S.71)

Einleitend wurde argumentiert, daß der experimentellen Psychologie in diesem Gegenstandsbereich eine entscheidende Rolle zukomme. Das trifft insbesondere für die gegenwärtige Situation zu (wie sie oben charakterisiert wurde). Die Vielfalt der bei der Analyse bereits gewonnenen experimentellen Ergebnisse kann heute nicht mehr - zugunsten etwa einer globalen Projektionshypothese - übergangen werden und ist auf nicht-experimentelle Weise weder zu widerlegen, noch zu vereindeutigen. Insbesondere die Bestimmung der Gültigkeit eines Verfahrens, also seiner diagnostischen Aussagemöglichkeiten, kann nur anhand weiterer experimenteller Ergebnisse vorbereitet werden [1].

Das Experiment spielt daher eine entscheidende instrumentelle Rolle für die Theorie, wie für deren Anwendung als Interpretationsgrundlage. Doch ist mit einer Entscheidung für den Weg der experimentellen Analyse (im Gegensatz zu einer allein erfahrungsgeleiteten Interpretation) nicht

---

1) Das wird bei der Diskussion des Begriffs der 'differentiellen Validität' noch ausführlicher zu erläutern sein.

unmittelbar eine Vorentscheidung getroffen über die Bewertung der anderen oben genannten psychologischen Richtungen.

Gleichwohl macht die experimentelle Psychologie gegenwärtig eine 'einzelheitliche Vorgehensweise' notwendig; ganzheitliche Interpretationen müssen vorerst als - jedenfalls e m p i r i s c h - unbegründbar zurückgestellt werden [1]. Wohl aber sind die Implikationen experimentellen Vorgehens unvereinbar mit einer geisteswissenschaftlichen Psychologie; das 'Verstehen' oder 'Deuten' (Heiß, z.B. 1951) der 'Zeichengehalte' eines Testprotokolls negiert - zugunsten des Evidenzcharakters der sprachlichen Inhalte von Testantworten - die Vielfalt der Bedingungen für das Zustandekommen der Reaktionen, - eben jene Vielfalt, welche durch die experimentelle Analyse systematisch erfaßt werden soll [2]. Offener ist das Verhältnis zur Psychoanalyse; denn hier sind theoretische Konzeptionen gegeben, welche das Zustandekommen einer Testreaktion bzw. deren Bedingungen genetisch begründen und daher auch zu Leitgedanken für experimentelle Analysen werden können [3].

Die experimentelle Forschung läuft jedoch stets Gefahr, den Gegenstand ihrer Untersuchung allein nach den verfügbaren Methoden zu bestimmen. Dadurch muß - angesichts der Vielfalt der heute schon verwendeten Ansatzpunkte für eine experimentelle Analyse - für den Kliniker das Bild eines

---

1) Die Versuche, 'klinische' Interpretationen zu validieren, indem ein 'ganzheitliches' Gutachten und somit die Einheit von Test und Diagnostiker zur Grundlage von verschiedenen Zuordnungsverfahren gemacht wird, können nicht als 'experimentell', weil in dieser einmaligen Kombination nicht wiederholbar, verstanden werden (vgl.S.).

2) Der Begriff der 'Mehrfunktionalität des Zeichencharakters' (Heiß 1951) beschreibt zwar diese Vielfalt, verschleiert sie aber, indem gerade durch diese Mehrdeutigkeit die Methode der 'Deutung' gerechtfertigt wird und eine systematische Analyse nicht als notwendig erscheint.

3) vgl. die Abhandlung der verschiedenen Projektionsbegriffe S. 14ff

atomistischen Vorgehens entstehen, welches zwar allgemeinpsychologisch interessierende Einzelaspekte untersucht, nicht jedoch die Diagnose von 'lebendigen Individuen' zu begründen vermag. Wieweit sich eines Tages auch ein stärker idiographisches Vorgehen empirisch begründen lassen mag, kann gegenwärtig nicht entschieden werden. Indessen ist es heute unzweifelhaft, daß die experimentell kontrollierte Testinterpretation in ihrer Treffsicherheit dem allein erfahrungsgeleiteten idiographischen Vorgehen überlegen ist. [1)]

Allerdings trifft die Kritik, die experimentelle Analyse sei atomistisch, zu, solange nur die in der a l l g e m e i n e n Psychologie bereits erprobten Forschungsmethoden auf die Analyse 'projektiver stimuli' ausgedehnt werden und solange nicht Theorien konzipiert werden, welche die Aussagemöglichkeiten eines projektiven Verfahrens in den Begriffen einer P e r s ö n l i c h k e i t s theorie präzisieren. Denn ein solches persönlichkeitstheoretisches Konzept würde n e u e Untersuchungsmethoden erfordern, welche dem Gegenstand 'projektiver Test' u.U. angemessener sind als jene allgemeinpsychologischen Methoden. Es ist nicht zu kritisieren, d a ß allgemeinpsychologische Modelle zur Analyse 'projektiver' Reaktionen herangezogen werden. Zu fragen aber ist, ob auf diesem Wege a l l e i n gerade jene Information in ausreichendem Maße gewonnen werden kann, welche den Test für die Diagnostik, insbesondere die klinische Diagnostik erst brauchbar macht. [2)]

Der historische Umstand, daß heute im Rahmen der experimentellen Forschung die in der allgemeinen Psychologie bereits bewährten Techniken und damit vorwiegend auch allgemeinpsychologische F r a g e s t e l l u n g e n übertragen

1) Meehl (1954) analysierte 20 Untersuchungen, in welchen die Erfolge der experimentell-statistischen Vorgehensweise und der klinischen Interpretation miteinander verglichen werden : erstere ist in ihrem Vorhersagewert der klinischen Beurteilung entweder gleich oder aber überlegen, nie jedoch unterlegen.

2) vgl. dazu Fußnote 1) auf S. 82

werden auf den Problembereich 'projektive Tests', darf nicht dahingehend falsch interpretiert werden, als ließen projektive Verfahren sich (wenn überhaupt) nur ebenso 'eindimensional' auswerten wie psychometrische Tests. Es darf nicht vorschnell daraus gefolgert werden, projektive Tests ließen beispielsweise nur die 'Messung' einzelner isolierter Motive zu und lieferten keine Hinweise auf individuelle Erlebnistatbestände, Konflikte, besondere Angstbesetzungen u.s.w. [1)]

G e g e n w ä r t i g kann jedoch die experimentelle Forschung nur einzelheitlich vorgehen. Noch müssen Ansatzpunkte für eine sowohl methodisch angemessene, wie theoretisch aufschlußreiche Analyse erprobt werden. Es werden daher in dieser Arbeit zunächst einzelne bisher gewonnene Konzeptionen in ihrem Ertrag zu beurteilen sein, bevor ein eigener Ansatz hergeleitet werden kann. An dieser Stelle sei auf das sehr ausführliche und übersichtliche Sammelreferat von Hörmann (1964a) aufmerksam gemacht und eine ähnliche Übersicht über den gesamten Bestand experimenteller Ansatzpunkte nicht eigens wiederholt.

In diesem Sammelreferat von Hörmann aber ist besonders auf einen Punkt hinzuweisen, der schließlich zur Anlage der vorliegenden Arbeit geführt hat. Hörmann greift einen Gesichtspunkt auf, der erstmalig von Rosenzweig (1950) formuliert wurde und seither in dieser Form nicht wieder in der Literatur behandelt wurde. Es ist dies die Frage nach den 'Funktionsebenen' eines Tests, die nun näher zu erläutern ist.

Konnte man bei der Annahme von 'klassischer Projektion' noch gewiß sein, der Proband würde insbesondere verdrängte und ich-bedrohliche Inhalte den Objekten seiner Fantasie-

---

1) Auf das heute noch unlösbare Problem, daß nur auf isolierten Merkmalsklassen basierende Wahrscheinlichkeitsaussagen sich experimentell sichern lassen, aber der klinischen Aufgabenstellung allein idiographisch-ganzheitliche Diagnosen gerecht werden können, wird am Schluß noch einmal zurückzukommen sein.

Produktionen (vor allem beim TAT) zuschreiben, so mußte man - wie bereits dargelegt - diese Annahme bald zu der allgemeinen Formulierung erweitern, der Proband würde Innerseelisches oder Ich-Spezifisches im weitesten Sinne in das Material hinausverlegen. Es wird nun zugestanden, daß auch bewußte, ja auch 'angenehme' Eigenschaften im Testmaterial zum Ausdruck gelangen können,und schließlich wird auch Vorstellungen der Eingang in solche Tests eingeräumt (vgl.S 9). Damit aber stellt sich als neue Frage : handelt es sich bei einer manifesten Antwort um Verdrängtes, um ein positiv bewertetes, also akzeptiertes Bedürfnis, um ein negatives, aber bewußtes Bedürfnis, um eine Vorstellung, eine Meinung, also eine Kognition und so z.B. um eine nicht-realisierte Ich-Ideal-Vorstellung?

Hörmann erläutert diese verschiedenen Möglichkeiten am Beispiel einer Reaktion auf die Wunschprobe von Wilde (1950), nämlich an folgender Begründung einer positiven Antwort auf das item 'Tiger' : 'weil der so schön beißen kann'. [1] Die sich in dieser Antwort sprachlich manifestierende Aggression läßt eine Reihe von Interpretationsmöglichkeiten zu :

> "daß die Vp aggressiv ist und aggressiv sein will, oder daß sie nicht aggressiv ist, aber aggressiv sein möchte, oder daß sie nicht aggressiv ist und auch bewußt nicht aggressiv sein möchte, ihre unbewußten Wünsche nach Aggressiv-sein aber hier projiziert, oder daß sie aggressiv ist, aber bewußt nicht aggressiv sein möchte und deshalb hier ihr Aggressiv-Sein-Wollen projiziert."(S.94)

---

1) In der Wunschprobe werden dem Pb 66 Karten vorgelegt, auf welchen je ein Wort wie etwa 'Tiger', 'Grashalm', 'Flugzeug' abgedruckt ist. Der Pb erhält folgende Instruktion : er solle sich vorstellen, daß er diese verschiedenen Gegenstände selbst s e i n könne; er solle bei allen 66 Karten wählen, ob er der betreffende Gegenstand sein möchte oder nicht. Dazu gibt es noch die weitere Kategorie, daß man etwas sowohl sein möchte, wie aber aus anderen Gründen zugleich ablehnen muß. Diese Wahlen sind im nachhinein durch den Pb zu begründen. Für weitere Angaben zur Wunschprobe siehe S.29f und S.125ff

Daß in dieser Testantwort [1] in irgendeiner Weise 'Aggression' angesprochen ist, ergibt sich direkt aus dem sprachlichen Inhalt : das wird um so mehr deutlich, wenn man ihn einer Antwort wie 'weil er ein schönes Fell hat' oder 'weil er frei ist' gegenüberstellt. Doch in welchem Sinne ist die Aggression hier zu verstehen? Was indiziert eine solche Antwort unter vielen anderen denkbaren Antworten zum gleichen item? Ein eindeutiger Schluß, eine Punkt-zu-Punkt-Zuordnung von Index und Indiziertem ist ganz offensichtlich von vorneherein und unbesehen nicht möglich. Eine Orientierung am Projektionsbegriff führt vollends nicht weiter : denn mit welcher Art von Projektion (klassisch, attributiv etc.) mag die Vp reagiert haben? Selbst die einfache Frage, ob die Vp aggressiv ist oder nicht, läßt sich nicht nach dem manifesten Inhalt der Test-Reaktion (in der falschen Voraussetzung einer 'content-validity') beantworten.

Welche Lösungsmöglichkeiten wären denkbar? Im einfachsten Falle müßte man nachweisen können (und es müßte dies für jede Verwendung des Tests allererst nachgewiesen s e i n !), daß dieser Test - aus welchen Gründen auch immer - gerade die nicht-realisierten Bedürfnisse anspreche (vgl. die alternative channel-Hypothese, s.Kap. II). Es ist aber ebenso gut denkbar, daß dieser Test, insofern er einen Auslöser für irreale, hypothetisch-fantasiemäßige Wunschvorstellungen darstellt und darin wiederum der Kontrolle des Bewußtseins weitgehend entzogen sein könnte, daß dieser Test also insbesondere die unbewußten Wünsche anspricht und in einer für das Subjekt unbewußten Weise zur Darstellung bringt. Aber warum sollten es nicht gerade die verhaltensmäßig realisierten, doch zugleich sozial-tabuierten Motive sein, für welche sich hier - ohne zu befürchtende soziale Sanktionen - ein unproblematischer freier Ausdruckskanal anbietet? Und welche Gründe sprechen dagegen, daß eine Vp sich an ihrem eigenen bewußten Selbstbild orientiert, wenn sie die verschiedenen items unter der Aufforderung zur Identifikation wahrnimmt?

1) welche ausgeführt etwa hieße : i c h möchte ein Tiger s e i n , w e i l der so schön b e i ß e n kann

Sieht man sich schon vor diese verschiedenen Möglichkeiten des Verhältnisses von Index und Indiziertem gestellt, so folgt noch die weitere Frage : Liegen die Bedingungen für die eine oder andere Reaktionsform im Test allein oder zugleich auch im Probanden? Kann man dem Test eindeutig eine bestimmte Funktionsweise zuordnen, so daß man sich etwa durchgängig darauf verlassen könnte, er würde die unbefriedigten Bedürfnisse zu Tage fördern? Oder liegen alle diese verschiedenen Möglichkeiten gleichzeitig in diesem einen Verfahren beschlossen, so daß es an der Eigenart (welcher Eigenart?) des Probanden läge, auf die eine oder andere Weise zu reagieren? Kann man dann aber vermuten, daß ein bestimmter Proband durchgängig im Sinne einer bestimmten Reaktionsform auf die Wunschprobe antwortet?

Daran schließen sich wieder zwei weitere Denkmöglichkeiten an : reagiert ein bestimmtes Individuum durchgängig bei allen projektiven Tests auf einer bestimmten 'Ebene' oder aber ist es bereits schon ein einziger Test (also hier die Wunschprobe), welcher eine einzelne Vp zu ganz verschiedenen Funktionsweisen stimuliert? Besteht vielleicht sogar ein diagnostisch wichtiger inter-individueller Unterschied gerade darin, daß die eine Vp konsistent im Sinne einer Reaktionsform den Test beantwortet, eine andere Vp aber inkonsistent reagiert?

Es ist dies das Problem der 'Funktionsebene', wie Rosenzweig es gesehen und 1950 formuliert hat : er vermutet nicht nur, daß ein bestimmter Test bei verschiedenen Vpn je verschiedene Reaktionsformen hervorruft (was er im übrigen am Beispiel von PFT und TAT erläutert), sondern, daß auch eine einzelne Vp in nicht konsistenter Weise auf einen einzelnen Test reagiert, daß also ihre Antworten ganz verschiedene Relationen zum Subjekt repräsentieren .

Das hieße beispielsweise in Hinsicht auf die Wunschprobe : eine Vp reagiert auf das item 'Tiger' mit einem unbewußten Aggressions-Wunsch; auf das item 'Rose' mit dem bewußten Bedürfnis, bewundert zu werden; auf das item 'Grashalm' mit der in ihrem Verhalten offen zu Tage liegenden Ängstlich-

keit; auf das item 'Orkan' mit der bewußten und verhaltensrelevanten Ablehnung zerstörerischer Aggressivität; auf das item 'Gebirge' mit dem realiter vorhandenen, bewußt abgelehnten, aber unbewußt wirksamen Bedürfnis nach Dominanz.

Daß diese Erwartungen, wie sie hier i.b. auf die Wunschprobe formuliert wurden, sich nicht mit jenen decken, welche Wilde bei seiner ersten (und leider einzigen [1]) Veröffentlichung an dieses Verfahren knüpfte, ist unmittelbar evident. Doch steht - allgemein gesehen - der hier zu entwickelnde Ansatz nicht notwendig in Widerspruch zu den bisherigen Annahmen über einen Test, sondern versucht, diese zu differenzieren.

Wilde geht einmal unkritisch von der Projektionshypothese aus, von der Annahme, daß "man Wünsche und Tendenzen des eigenen So-seins auf fremde Wesen und Dinge projiziert" (S.215) [2]. Zum anderen aber fehlt vor allem eine Differenzierung dessen, was 'projiziert' wird. Was sind 'die Dimensionen der charakterologischen Dynamik, das Relief der persönlichen Tendenzen, die Spannungen, die So-Seins-Wünsche, die Richtungen" (S.215), welche in den Zuordnungen und Begründungen zum Ausdruck kommen sollen? Ihre theoretische Herkunft ist nicht ersichtlich und der diagnostische Wert solcher Angaben ob ihrer Allgemeinheit gering. Es fehlt eine Persönlichkeitstheorie, welche die Rolle solcher Spannungen und Richtungen allererst begrifflich definieren und inhaltlich bestimmen könnte. Das oben angeführte Beispiel möglicher Interpretationen einer Aggressions-Antwort zeigt, daß mit dem rein deskriptiven Begriff einer 'Tendenz' eher der Zugang zu dem indizierten Phänomen versperrt wird, als

---

1) Wilde starb, bevor er die in Aussicht gestellten Untersuchungen zu seinem Verfahren veröffentlichen konnte.

2) Diese Annahme ist hier allerdings u.U. eher gerechtfertigt als z.B. beim TAT. Es wird ja bei der Wunschprobe eine Identifikation direkt gefordert: man soll sich vorstellen, ob man die verschiedenen Gegenstände selbst sein möchte, während beim TAT nicht die analoge Aufforderung explizit ausgesprochen wird, der Pb solle sich vorstellen, eine der abgebildeten Personen selbst darzustellen.

daß es bereits in irgend hinreichender Weise gefaßt oder gar 'erklärt' wäre. So wird man gerade nicht die "Soseins-Tendenz des Probanden g a n z u n m i t t e l b a r [1] ablesen können" (S.224). Sondern man wird vielmehr die Antworten sehr präzise differenzieren müssen nach der Vielfalt möglicher indizierter psychischer Tatbestände. Die Tatsache einer irgendwie festgestellten Aggressions-'Tendenz' ist in sich noch kein diagnostischer Aufschluß, sondern stellt erst ein diagnostisches Problem : ist es eine bewußte oder unbewußte, eine befriedigte oder unbefriedigte, d.h. also eine realisierte Tendenz oder eine solche, die vermittels Aggressions-Angst vom Verhalten ferngehalten wird?

Es ist dies die allgemeine Problemstellung, welche dieser Arbeit vorausgeht. Sie entbehrt bis heute einer theoretischen Systematisierung, wie empirisch gesicherter Grundlagen, von welchen bereits experimentelle Hypothesen deduzierbar wären. Vielmehr gilt es nachzuweisen, daß diese Frage nach den Funktionsebenen eines Tests weiterführend, also heuristisch fruchtbar ist.

Dieser Nachweis soll im experimentellen Teil am Beispiel der Wunschprobe versucht werden. Es wird an späterer Stelle noch aufgezeigt werden, warum dieses Verfahren sich in besonderer Weise für eine solche Fragestellung anbietet. Es fungiert daher die Wunschprobe hier als B e i s p i e l : es wird erwartet, daß diese Fragestellung auch für andere projektive Verfahren relevant ist (vgl. Rosenzweig 1950); doch können Ergebnisse der Wunschprobe nicht direkt auf andere Tests übertragen werden. Zugleich wird nicht der Anspruch erhoben, es würde die Wunschprobe ihrerseits irgend erschöpfend behandelt : sie bietet eine Vielfalt anderer Probleme und wahrscheinlich auch weiterer diagnostischer Möglichkeiten, als hier angesprochen werden können. Die Wunschprobe dient hier vielmehr als M i t t e l zur Erprobung eines bestimmten theoretischen Ansatzes, nämlich dem der 'Funktionsebenen', und ist selbst nicht Z w e c k , etwa im Sinne eines Validierungsvorhabens.

---

1) gesperrt durch Verfasser

Die Annahme, daß e i n projektiver Test 'auf verschiedenen Ebenen" funktionieren könne, gründet bis heute allein in der E r f a h r u n g im Umgang mit solchen Tests (so bei Rosenzweig 1950). Es ist daher zunächst zu prüfen, wieweit diese Annahme durch die Ergebnisse der experimentellen Erforschung projektiver Methoden bestätigt wird. Es ist also zu untersuchen, ob die bis heute experimentell nachgewiesenen Formen des Testverhaltens bei projektiven Tests eine begriffliche Ordnung mit Hilfe des Ebenen-Konzeptes erlauben oder gar notwendig machen. Erst bei einer solchen Analyse vorliegender Experimente können sich auch die Gesichtspunkte ergeben, nach welchen ein Ebenen-Modell, das sich für die e x p e r i m e n t e l l e Bestimmung einzelner Funktionsebenen eignen soll, neu konzipiert werden kann.

Ist ein solches Ebenen-Konzept in Orientierung an derartigen experimentellen Ergebnissen gewonnen worden, so wird auch dieses noch nicht einen theoretischen Rahmen bieten können, aus dem sich exakt Hypothesen deduzieren lassen. Zunächst wird sich nur die Erwartung p r ä z i s i e r e n lassen, daß die Reaktionen auf projektive stimuli bestimmten 'Ebenen' zugeordnet werden können. Eben diese Erwartung ist einem Erkundungsversuch zu unterziehen. Erst einem solchen können sich dann gezielte Entscheidungsversuche anschließen.

3. Vorausschau auf den Gang der Darstellung

Von obiger Aufgabenstellung her ergibt sich nun folgender Gang der Arbeit.

Es sind allererst die Grundlagen zu schaffen, von welchen aus ein Ebenen-Konzept entwickelt werden kann (ERSTER TEIL).

Diese Grundlagen sind vorwiegend aus der Literatur zu gewinnen : durch die kritische Betrachtung aller bisherigen Formulierungen eines Ebenen-Konzeptes einerseits (Kap.I) und andererseits durch eine Analyse der bisherigen empirischen Untersuchungen daraufhin, wieweit die vorliegenden

experimentellen Ergebnisse, die im Rahmen anderer theoretischer Konzepte gewonnen wurden, eine Betrachtung im Rahmen eines Ebenen-Modells zulassen bzw. mehr noch erfordern (Kap. II).

Eine solche Übersicht nun leistet ein weiteres : sie macht deutlich, mit welch erheblichen Problemen eine Validierung dieser Verfahren verbunden ist; sie verweist aber zugleich darauf, daß sich systematische Relationen zwischen Index und Indiziertem überhaupt feststellen lassen, so daß eine Validierung einzelner Test- s c o r e s , anstelle einer Validierung der durch den Interpreten gewonnenen D i a g n o s e , nicht mehr als ausgeschlossen gelten muß. Jede Konzeption, welche für die Analyse projektiver Verfahren neu eingeführt wird, hat sich daher insbesondere daran zu bewähren, ob sie Ordnungsgesichtspunkte liefert, welche als Grundlage für künftige Validierungsuntersuchungen dienen können. Auch diese Frage ist also im vorhinein theoretisch zu klären, bevor es als hinreichend gerechtfertigt erscheinen kann, ein Ebenen-Konzept für die Analyse projektiver Methoden zu entwickeln (ZWEITER TEIL). Läßt sich also zeigen, daß ein Ebenen-Konzept jene Bedingungen erfüllt, welche es [redacted] als einen Rahmen für Validitätsuntersuchungen qualifizieren (Kap. III), und ist sodann das theoretische Modell selbst entwickelt worden (Kap. IV), so können schließlich auch die experimentellen Bestimmungen hergeleitet werden, welche eine Erkundung des Ebenen-Konzeptes realisieren (Kap. V).

Da hier für den Zugang zu den Persönlichkeitsebenen ein in Deutschland offenbar noch ungebräuchliches Verfahren benutzt wurde, nämlich die 'Q-sort-Methode', ist diese eingehender dargestellt worden. Sie wird im Text jedoch zunächst nur so weit allgemein entwickelt, wie es für ein Verständnis der Versuchsanordnung unbedingt erforderlich ist. Die methodologischen Implikationen und Probleme sind demgegenüber in einen Exkurs verlegt, da sie zwar als implizite Voraussetzungen des Experiments einer theoretischen Klärung bedürfen, jedoch in direkter Weise nicht zur eigentlichen Planung des Experiments beitragen (Kap. VI).

Sind somit der theoretische Standort und die experimentelle Fragestellung im einzelnen entwickelt, so kann nunmehr das Experiment selbst dargestellt werden (DRITTER TEIL). Es wird zunächst die Durchführung des Experiments und die Art des Datenmaterials näher zu charakterisieren sein (Kap. VII), und es sind schließlich die Ergebnisse im einzelnen zu interpretieren (Kap. VIII). Dabei muß es sich teilweise um eine Interpretation post festum handeln, da - wie bereits erläutert wurde - wohl leitende Gesichtspunkte, nicht jedoch exakte Hypothesen dem Erkundungs-Experiment vorangestellt werden konnten.

Im Schlußteil wird dann der ganze Ansatz noch einmal kritisch zu durchleuchten sein (Kap. IX).

E R S T E R T E I L

Das Konzept der Funktionsebenen in seiner Relation zu anderen Forschungsansätzen in bezug auf projektive Tests

I. Das Problem der Funktionsebenen, soweit es in der Literatur explizit formuliert wird.

1. Das Ebenen-Problem in seiner ersten Fassung: ROSENZWEIG

Es war Rosenzweig, der 1950 als erster das Problem der Funktionsebenen formuliert hat und zwar sowohl in seiner allgemeinen Relevanz für die diagnostischen Testmethoden überhaupt, wie aber vor allem in seiner speziellen methodologischen Bedeutung für die projektiven Testverfahren.

Diesem Begriff der Funktionsebene eines T e s t s korrespondiert der Begriff der Ebene des V e r h a l t e n s. Menschliches Verhalten selbst spielt sich, nach Rosenzweig, auf verschiedenen Ebenen ab, d.h. : diese 'Ebenen' differenzieren das Verhalten begrifflich nach Inhalt, Äußerungsform und psychologischer Bedeutung. Diesen verschiedenen Ebenen kann auch das Verhalten in diagnostischen Situationen, insbesondere auf Tests, zugeordnet werden. Der Begriff 'Ebene' ist also ein persönlichkeitstheoretisches Konzept, welches zur Analyse diagnostischer Methoden herangezogen wird. Indessen handelt es sich nicht um eine 'Theorie', sondern um ein deskriptives Modell, dessen theoretische wie empirische Bewährung noch ausstehen.

Rosenzweig unterscheidet 3 Ebenen und nennt das durch sie bestimmte Verhalten : opinion behavior, overt behavior und implicit behavior. Sie unterscheiden sich der angegebenen Reihenfolge nach hinsichtlich ihrer 'Tiefe' bzw. umgekehrt hinsichtlich des jeweils charakteristischen Grades an kognitiver Zensur. Rosenzweig definiert :

Level I = opinion behavior : "The subject functions according to a rather deliberately invoked standard of what he considers to be right or proper, appropriate or otherwise acceptable to himself or others in the particular situation and, if necessary, he may distort his original reaction tendencies to fit these opinions." (S.64)

Level II =overt behavior : "What emerges here is a sample of the subject's usual observable performences under external conditions defined in common sense terms." (S.65)

Level III=implicit behavior : "...includes reactions which in their full significance are usually denied expression but which achieve some partial or symbolic manifestation under favorable conditions." (S.65)

Diese 3 Ebenen lassen sich zwar begrifflich klar scheiden. Doch die psychischen Vorgänge selbst sind eher durch eine Überlappung und Vermischung der 3 Verhaltensmodi gekennzeichnet. Es können die Ebenen nach Rosenzweig sogar soweit zusammenfallen, daß Unterschiede zwischen ihnen nicht aufgedeckt werden können : "... the levels are apt to mingle and merge kaleidoskopically to a degree that defies analysis." (S.65)

In jedem Falle aber setzt die Beurteilung einer Person auf einer 'Ebene' das Wissen über ihr Verhalten auf den anderen Ebenen voraus : erst die gleichzeitige Analyse aller 3 Ebenen ermöglicht, daß jede einzelne Ebene angemessen, nämlich im Lichte der übrigen, beurteilt wird. Meinungen über das eigene Verhalten oder auch über Verhaltensstandards (level I) gewinnen erst dann eine diagnostische Bedeutung, wenn sie verglichen werden können mit dem tatsächlichen Verhalten (level II). Objektive Verhaltensdaten sind für den Diagnostiker erst dann informativ, wenn er sie beziehen kann auf das Selbstbild der Person und auf ihre Verhaltensstandards, und wenn er dieses manifeste Verhalten vergleichen kann mit jenen Inhalten, welche die Person nach außen hin verbirgt (level III). Jedoch handelt es sich bei dieser Aufgliederung um nicht mehr als um allgemeine Annahmen, die einer näheren Bestimmung noch bedürfen : "Any comprehensive approach to personality must include them all (all levels, Verf.). The essential problem hence becomes

that of determining the principles according to which the levels influence each other, now by opposion and now in cooperation." (S.66)

Der Begriff der Ebene dient jedoch nicht allein der Differenzierung verschiedener Verhaltensaspekte. Er führt darüber hinaus zu einer ganzheitlich-integrativen Betrachtung der Person : "It thus becomes essential for the making of definite interpretations not only to know at which level the subject has performed but also to have information about the integration of his personality in terms of these very levels." (S.69)

Für die Frage, in welcher Weise projektive Tests einen Zugang zu diesen Persönlichkeitsebenen erschließen, ist - darauf war einleitend schon eingegangen worden - vor allem folgende These ausschlaggebend : bestimmte diagnostische Verfahren garantieren nicht jeweils bestimmte Reaktionsformen (sprich 'Ebenen'),und insbesondere die projektiven Methoden rufen nicht - wie in der Projektions-Hypothese angenommen wird - notwendig nur latente Inhalte (Ebene III) hervor. Vielmehr können projektive stimuli nicht minder Reaktionen im Sinne der Ebenen I und II herbeiführen, was Rosenzweig überzeugend am Beispiel von PFT und TAT veranschaulicht.

Es wird also bei Rosenzweig die Erwartung, wenn auch nicht experimentell bestätigt, so doch überzeugend anhand von Beispielen abgeleitet, daß ein bestimmter projektiver Test nicht eindeutig auf einer und nur einer bestimmten Ebene funktioniere. Weiter nimmt Rosenzweig an, daß sogar - über diese Variation eines Tests von Person zu Person hinaus - auch das Verhalten e i n e r Person in bezug auf e i n e n Test inkonsistent sein könne. Damit ist zwar offengelassen, ob nicht ein bestimmter Test zumindest vornehmlich nur Reaktionen e i n e r bestimmten 'Ebene' hervorrufe; aber es bleibt doch bei jeder Versuchsperson neu zu analysieren, ob sie tatsächlich auf dieser Ebene reagiert und wieweit sie es in konsistenter Weise tut.

Diese genaue Bestimmung der tatsächlichen Reaktionsweise im einzelnen Fall ist nicht ohne zusätzliche Information möglich. Die Vorschläge allerdings, die Rosenzweig zu diesem Punkt gibt, sind wenig befriedigend. Eine eingehende Exploration (auf welcher Basis man seine Antworten gegeben habe, ob man sich selbst auf (z.B.) den TAT-Bildern habe sprechen lassen) soll die jeweilige Funktionsebene bestimmen helfen. Varianten jedoch, die zunächst vom Verhaltensaspekt her definiert wurden (also die 'Ebenen'), können kaum vom Erlebensaspekt her allein beurteilt werden und dies insbesondere nicht unter den Bedingungen einer diagnostischen Situation. Gerade hier kann (vielleicht im Unterschied zu einer experimentellen Situation, in der sich die Einstellungen und Motivationen der Vpn u.U. manipulieren lassen) am wenigsten der indirekte Zugang über Testverfahren durch direkte Explorationsmethoden ersetzt, wie auch schwerlich nur ergänzt werden.

2. Das Ebenen-Konzept als Hilfskonstruktion zur Interpretation fehlender Validität bei projektiven Tests

Bei allen übrigen Autoren, die den Terminus 'level' bei der Analyse der Funktionsweise projektiver Tests verwenden, wird obige Arbeit von Rosenzweig nicht erwähnt. Der Begriff findet sich hier vielmehr nur in sehr reduzierter Form wieder, so etwa bei Saxe (1950), Gluck (1955) und Davids und Rosenblatt (1958) [1]. Gemeinsam ist diesen Autoren, daß der Begriff 'level' ohne theoretischen Kontext, wie ohne experimentelle Realisation herangezogen wird, um negative Ergebnisse bei Validierungsuntersuchungen und das Fehlen von signifikanten Korrelationen zwischen verschiedenen projektiven Tests einleuchtend zu machen. Die Differenz von Testverhalten und overt-Verhalten, von Testmaterial und klinischem Material und die unterschiedlichen Ergebnisse im Vergleich mehrerer projektiver Tests untereinander werden hier

1) Davids/Rosenblatt beziehen sich auf eine Reihe anderer Autoren, die jedoch ungenannt bleiben.

rückwirkend ihrer - im Sinne der Validierungsabsicht - negativen Bedeutung entkleidet. Diese unterschiedlichen und scheinbar widersprüchlichen Ergebnisse finden für die Autoren eine hinlängliche Erklärung in der Annahme, daß verschiedene Tests (auf nicht näher bestimmte Weise) Verschiedenes messen : man müsse zwischen 'Oberfläche' und 'Tiefe', zwischen bewußter und unbewußter Motivation, zwischen verschiedenen Verteidigungsmechanismen und zwischen verschiedenen Graden der 'social desirability' [1] bei den Persönlichkeitsvariablen unterscheiden. Damit allerdings ist das Problem erst benannt, nicht jedoch schon geklärt.

Der Begriff der 'Ebene' hat in dieser Diskussion eine doppelte Funktion. Zum einen hebt das Konzept verschieden starker Bewußtseinskontrolle (sprich 'Ebenen') bei Testreaktionen die scheinbaren Widersprüche in den Testergebnissen durch den Hinweis auf unterschiedliche Bedingungen auf (s.o.). Zum anderen wird die als unzureichend empfundene Dichotomie zwischen 'bewußt - unbewußt', 'oberflächlich - tief', 'overt - covert' ersetzt durch ein K o n t i n u u m der 'Tiefe' bzw. 'Zugänglichkeit', welches freilich mehr als derartige Dichotomien jeder Präzisierung entbehrt. Es fehlt - im Gegensatz zu Rosenzweig - jede nähere Bestimmung der Ebenen selbst. Es handelt sich eher um ein Betrachtens-Schema, denn um ein theoretisches Konzept : "We may picture personality as containing different layers or strata, varying in their degree of accessibility to observation from outside and to self observation, and also in acceptability to the person himself" (Hanfman/Getzels 1953, S.281).

Auch der neuere Versuch, das Konzept der Ebenen auf den psychologischen Mechanismus von Antrieb und Kontrolle zurückzuführen, ist - so jedenfalls bei Stone und Dellis (1960) [2] - nicht mehr als eine Setzung, deren theoretische Implikationen ungeprüft bleiben und aus denen experimentelle Hypothesen nicht abgeleitet werden : "Our general frame of reference is that personality is organized at

1) für eine kritische Analyse dieses Begriffs siehe S.179ff
2) vgl. dazu S. 82

different levels of impulse-control systems. We believe that these different levels of personality organization consist of a variety of impulse-control-systems which exist on a continuum from those which are very primitive and undifferentiated to those which are highly organized and differentiated" (S.336). Es fehlt jede Angabe darüber, wie die jeweilige Organisation eines impulse-control-systems zu denken sei. Demzufolge kann nur vage als Hypothese formuliert werden,"that the appearance of a particular trait in one impulse-control-system may have an entirely different significance from it's appearance in another impulse-control-system" (S.338). Auch die Art dieses Unterschiedes wird nicht näher bestimmt, wennzwar erstmalig Annahmen über deren Bedingungen eingeführt werden. Der Test selbst nämlich wird als Auslöser für die jeweilige Persönlichkeitsebene (bzw. das impulse-control-system) verstanden. Das Ebenen-Konzept wird daher hier durch die Annahme einer systematischen Beziehung zwischen Test-stimulus und Reaktion begründet : je stärker die stimulus-Struktur, eine desto weniger tiefe Ebene wird durch den Test erreicht und umgekehrt (vgl. auch Carr 1956, Theiner 1962). Daß eine solche einfache Beziehung jedoch nicht besteht, wurde von anderer Seite vielfach nachgewiesen, so bei den Untersuchungen, die sich methodisch direkt auf das stimulus-Material projektiver Tests richten [1].

Bei aller Unzulänglichkeit der bisherigen Arbeiten zur Ebenen-Hypothese ist jedoch - das wurde schon in der Einleitung deutlich - mit Rosenzweig anzunehmen, daß das Ebenen-Konzept ein vielversprechendes heuristisches Prinzip darstellt, - welches bis heute viel zu wenig ausgeschöpft wurde, um systematische Ordnungsgesichtspunkte in die Vielfalt heterogener Testreaktionen einzuführen. Allerdings sind methodisch neue Wege und theoretisch gründlichere Analysen vonnöten. Das bloße Zuordnen von Test-scores zu vage behaupteten, jedoch weder theoretisch, noch operational definierten 'Ebenen' hat sich als fruchtlos erwiesen. Es

---

1) vgl. Kagan 1956, 1959; Crandall 1955; Epstein und Smith 1956; Farberow 1959 u.a.

fehlt an einer Persönlichkeitstheorie, innerhalb welcher die 'Ebenen' exakt definiert werden könnten. Und es fehlt - das wird im folgenden deutlicher werden - an einer Theorie des Test-V e r h a l t e n s , welche es erst leisten kann, die möglichen Reaktionen auf bestimmte Test-stimuli zu differenzieren und - wichtiger noch - mit Hilfe der bisher erarbeiteten allgemeinpsychologischen Modelle zureichend zu formulieren.

Ohne eine solche Grundlage kann das Ebenen-Konzept nicht über die Projektionshypothese hinausführen, sondern teilt mit dieser seine theoretische Unzulänglichkeit, wie den Mangel an empirischer Bewährung. Die aus der Ebenen-Hypothese hergeleitete Annahme, daß relativ unstrukturiertes Material die Bewußtseinskontrolle außer Kraft setze und die 'tieferen', verborgeneren Inhalte zu Tage fördere, war ja gerade der unkritische Ausgangspunkt jener Autoren gewesen, welche die sgn. projektiven Tests durch die Projektionshypothese hinlänglich erklärt glaubten.

Soll aber der verschwommene Projektionsbegriff durch ein der Sache angemesseneres Konzept ersetzt werden, so sind an ein solches zumindest 2 Bedingungen zu knüpfen : Es muß die Testantworten in einem inhaltlichen, wie in einem 'genetischen' Sinne differenzieren. Die Aufgliederung der Testantworten nach verschiedenen 'Ebenen', damit die Differenzierung ihrer inhaltlich-diagnostischen Bedeutung, kann eine theoretische Grundlage nur dadurch erhalten, daß die Testleistungen sich nach der je besonderen Art ihres Z u s t a n d e k o m m e n s , also ihrer Genese, bestimmen und unterscheiden lassen. [1] Die verschieden starke Korrelation zwischen Test und Kriterium kann nicht als hinreichende Grundlage für den Nachweis dienen, daß durch den

---

1) In allen bisherigen Versuchen einer theoretischen Begründung der Funktionsweise proj. Tests war es allein die Annahme der 'klassischen' Projektion, welche einen solchen genetischen Begründungszusammenhang lieferte (vgl.S. 11).

Test Verschiedenes indiziert würde. [1] Erst wenn festgestellt ist, aufgrund welcher Bedingungen derart verschiedene Reaktionen zustande kommen, lassen sich die unterschiedlichen Testergebnisse in einen theoretischen Zusammenhang stellen. Erst dann wird es auch möglich sein, die Kriterien selbst zu differenzieren, mit welchen der Test nunmehr in einen systematischen Zusammenhang gebracht werden kann. [2]

Es kann also der Test-score nicht abgelöst von den B e d i n g u n g e n der Reaktion, die zu ihm führten, verstanden werden, soll die experimentelle Zuordnung von Testleistung und Kriterium in ihren theoretischen Konsequenzen über die Projektionshypothese oder einen vagen Ebenen-Begriff hinausführen. Das Konzept der 'Ebene' kann so gesehen nur fruchtbar sein, wenn es nicht verschiedene sgn. projektive Tests auf einem imaginären 'Tiefenkontinuum' anordnet, sondern vielmehr verschiedene ' R e a k t i o n s ' - Ebenen systematisch voneinander isoliert. Soll ein solches Ebenen-Konzept zugleich den Aufgaben einer klinischen Diagnostik angemessen sein, muß es innerhalb eines persönlichkeitstheoretischen Rahmens formuliert werden können. [3] Soll es einen weiteren Realitätsbereich umfassen als der Projektionsansatz, muß es zugleich jene weiteren Reaktionsformen theoretisch mit umschreiben können, welche bisher experimentell nachgewiesen worden sind. (vgl. S. 55ff)

Es wird daher im folgenden zunächst ein Ebenen-Konzept zu betrachten und in seinen Grenzen zu bestimmen sein, welches den Kernbegriff einer an den Aufgaben der Persönlichkeits-

---

1) Es kann damit genau so überzeugend aufgewiesen werden, daß der Test nicht immer mißt, was zu messen von ihm behauptet wird, also - seine unzureichende Validität.
2) vgl. den Begriff der 'differentiellen Validität', S. 87ff
3) vgl. S. 24f und Fußnote 1) auf S. 82

diagnostik ausgerichteten Persönlichkeitstheorie darstellt. Schließlich wird in einem eigenen Kapitel zu untersuchen sein, welche allgemeinpsychologischen Erkenntnisse über die Reaktionen auf projektive stimuli in der experimentellen Forschung gewonnen wurden. Erst von dieser Grundlage aus kann dann ein Ebenen-Konzept entwickelt werden, welches umfassend genug ist, um empirisch prüfbare Hypothesen zu erschließen. Diese Hypothesen müssen die Erwartung v e r s c h i e d e n e r Formen der Testreaktion präzisieren. Sie können dann an die Stelle der vagen Vermutungen über layers, stratas oder segments treten.

3. Das Ebenen-Konzept im Rahmen einer Persönlichkeitstheorie : LEARY

Die Forderung, daß das Konzept der 'Ebenen' innerhalb einer Persönlichkeitstheorie formuliert werden müsse, ist vorrangig darin begründet, daß allein innerhalb eines solchen Rahmens sich einzelne 'Ebenen' begrifflich exakt genug bestimmen lassen, um den vagen Begriff des 'Ebenen-Kontinuums' ersetzen zu können. Eine solche Persönlichkeitstheorie wurde auf psychoanalytischer Seite von Leary (1957) entwickelt. Bemerkenswerterweise wurde auch sie, ähnlich der Arbeit von Rosenzweig (1950), in der weiteren Ebenen-Diskussion kaum berücksichtigt [1], obwohl der Begriff hier nicht allein als Basis des Persönlichkeitsverständnisses fungiert, sondern zugleich auch einen Schlüsselbegriff für die projektive Testdiagnostik darstellt [2].

Ausgangspunkt dieser Theorie ist, "that the human being is a complex, multilevel pattern of conflicting motives and behavior". Das bedeutet für jede Beschreibung von Persönlichkeit, aufgrund welchen Materials auch immer : "Any statement about personality must indicate the level of personality to which it refers" (S.41). Der Begriff der Ebene wird

---

1) mit Ausnahme von Stone, Dellis 1960, Filmer-Bennet, Klopfer (1962)

2) Daß die Theorie gerade in diesem Punkt kritisch einzuschränken ist, kann erst im folgenden deutlich werden.

hier sogar soweit zum Angelpunkt für Persönlichkeitstheorie überhaupt, daß er, nach Leary, alle Theorien gleichermaßen bestimmen müsse : "Although the content of theories may vary, it is hard to conceive of a personality theory (in the post Freudian era) which does not deal with the problem of levels" (S.42).

Zusammenfassend gesagt, wird hier der beachtenswerte Versuch unternommen, in der Konzeptualisierung der je spezifischen Beziehungen der Ebenen zueinander einen Zugang zur Einheit der Person zu gewinnen, - dies in der strikten Ablehnung einer isolierten Analyse einzelner Variablen auf zudem noch heterogenen Maß-Dimensionen. Wieweit dieser Versuch, auf der - durchaus begrenzten - Basis des hier allen Ebenen gemeinsam zugrundegelegten Kategoriensystems von Eigenschaften [1], tatsächlich gelingt, bleibe hier dahingestellt. Es sei nur hervorgehoben, daß die Theorie selbst, abgesehen von bestimmten Prämissen [2], strikt an den Grenzen ihrer empirischen Realisation, d.h. an den Möglichkeiten der Diagnostik, ausgerichtet ist, und daß dementsprechend auch für jede einzelne Ebene angegeben ist, wie sie empirisch zu erheben sei [3].

Die Ebenen nun selbst, die Leary aufgestellt hat, sollen im folgenden kurz einzeln charakterisiert werden. Sie erfüllen (zum Teil) die Forderung nach exakter Definition, müssen aber daraufhin geprüft werden, ob sie auch der zweiten Forderung entsprechen, einen Zugang zu den Reaktionen auf projektive stimuli, also zum Testverhalten, zu liefern. Leary unterscheidet 5 Ebenen, von denen die 3 ersten denjenigen von Rosenzweig kongruent sind (vgl. S. 35) :

---

1) Leary gründet diese Theorie auf 16 Variablen, welche in ihrer Ausprägung auf den verschiedenen Ebenen verglichen werden.
2) z.B. die Angstreduktion als grundlegendes Prinzip allen Verhaltens : "The basic function of the individual's interpersonal behavior is basic to ward off survival anxiety" (S.202).
3) vgl. Fußnote 2) der Vorseite

I. level of public communication
II. level of conscious "
III. level of private "
IV. level of the unexpressed
V. level of values

ad level I :

"The events studied at this level are the overt interpersonal activities of the individual." "The basic unit.... is the interpersonal effect."

Es wird auf dieser Ebene also das menschliche Verhalten (das hier prinzipiell 'soziales' Verhalten ist) vom Gesichtspunkt des Beobachters her gesehen. Verhalten wird dabei so verstanden, daß es mittelbar und prinzipiell die Motive reflektiert. Das gilt auch und gerade für das individuell charakteristische Rollenverhalten und für die spontanen Automatismen, wie sie Tonfall, Gestik und äußere Erscheinung bestimmen. Verhalten auf Ebene I ist somit : "involuntary, automatic behavior of which the subject can or cannot be aware" (S.99). Es wird erhoben durch Beobachtung, Situationstests, interpersonal-checklist und Indizes des MMPI.

ad level II :

Es handelt sich hier um "..the individual's perceptions of himself and his world as he (consciously) reports them". "...it reflects how the subject chooses to present himself and his view of the world." (S.133)

Es sind dies also Inhalte des Bewußtseins, soweit das Individuum bereit ist, sie auf dem Wege des self-report zu kommunizieren. Gegenüber allen nur denkbaren Selbstäußerungen sind als Einheiten dieser Ebene die 'interpersonellen' Eigenschaften bzw. die ihnen zugrundeliegenden Motive festgelegt. Die fehlende Zuverlässigkeit solcher Selbstaussagen wird dadurch eingeschränkt, daß sie zum einen anhand verschiedener Methoden und damit verschiedener Auslöserreize für die Selbstdarstellung, zum anderen in Situationen mit verschiedenem sozialen Druck und damit verschiedener Moti-

vation und auch zu verschiedenen Zeitpunkten erhoben werden. Je übereinstimmender die Aussagen, desto zuverlässiger die über Ebene II gewonnenen Daten.

ad level III :

"... level III comprises the expressions that an individual makes, not directly about his real self in his real world, but indirectly about an imagined self in his preconscious or symbolic world" (S. 154).

'Preconscious' bedeutet hier nicht das 'Vorbewußte' als 'System' im psychoanalytischen Sinne; ebenso bedeutete auf Ebene I 'bewußte' Selbstdarstellung nicht, daß das 'Bewußtsein' als solches zum Ausdruck gelange. Es handelt sich auch bei dieser Ebene vielmehr um "those themes which the subject is willing to express in the testing situation" (S.155) bzw. auch allgemeiner um den Inhalt von Träumen, Fantasie, schöpferischem Ausdruck und von Wünschen. Praktisch gesehen handelt es sich jedoch im wesentlichen um "responses to projective stimuli". Im Gegensatz zur Funktion der Realitätsanpassung, welche das Verhalten der Ebenen I und II charakterisiert, handelt es sich hier, in der 'symbolischen Repräsentation', um "responses to internal ambiguity and tension" (S.156). Die Funktion dieser symbolischen Vorgänge nämlich besteht darin, realiter ungelöste Probleme zu bewahren und Impulse aufzuschieben (time-binding), bis sie zu einem späteren Zeitpunkt gelöst bzw. befriedigt werden können. Daher können die symbolischen Inhalte nach Leary, insbesondere in ihrer Relation zu den Ebenen I und II, auch als Anzeichen dafür gewertet werden, ob eine Verhaltensänderung zu erwarten ist oder nicht. Dieses ist der Fall, wenn gegenwärtig latent vorliegende Probleme (Differenzen zwischen 2 Ebenen) es wahrscheinlich machen, daß sie zu einem anderen Zeitpunkt gelöst werden.

Fallen die Ebenen II und III zusammen, so liegen nach Leary entweder spezifisch 'private' Inhalte nicht vor; oder aber die Vermeidung eben dieser Inhalte ist so stark, daß sie selbst auf Ebene III nicht zum Ausdruck gelangen. Letzteres

wird insbesondere indiziert, wenn sich 'Fehlwahrnehmungen' nachweisen lassen oder die Vermeidung solcher Themen, welche durch den 'thematic-pull' der stimuli (z.B. des TAT) nahegelegt werden. [1] Wichtig ist bei dieser Kennzeichnung der Ebene III, daß die hier geäußerten Inhalte nicht ausschließlich als unkontrollierter 'vorbewußter' Ausdruck bewertet werden, sondern daß auch situative und damit motivationale Einflüsse bei diesen Fantasie-Reaktionen angenommen werden, daß also auf dieser Ebene eine Bewußtseinskontrolle nicht ausgeschlossen wird. Im Gegenteil ermöglicht erst der Vergleich der Inhalte von Ebene III mit jenen der Ebene II eine Entscheidung darüber, ob die 'private world' durch das verwendete Verfahren überhaupt erreicht wurde oder ob ein starkes Verteidigungsverhalten (resp. eine starke Angstbesetzung) eben dieses verhinderte. Nach Leary kann die 'private world' also verschieden 'tief' und u.U. schwer erreichbar sein.

Dieser Vergleich der Ebene III (also der Antworten auf projektives Material) mit den übrigen Ebenen liefert zweierlei Information : einmal kann (so meint Leary) entschieden werden, o b die private world zum Ausdruck gelangt; zum anderen wird die Bedeutung der Inhalte dieser 'private world' im Gefüge einer Person überhaupt erst durch eine solche Gegenüberstellung aller Ebenen erkennbar. Erst die Einsicht in das Verhältnis dieser Ebene zur Selbstwahrnehmung und zum tatsächlichen Verhalten macht eine diagnostische Aussage über die Person möglich. [2] Demnach kann ein projektiver Test also grundsätzlich nicht für sich genommen interpretiert werden : je nach dem Kontext der übrigen Ebenen indizieren die durch ihn hervorgebrachten Inhalte ganz verschiedene Gegeben-

---

1) Leary hat aufgrund klinischen Materials eigene Normen entwickelt; zur Fragwürdigkeit solcher Normen siehe S. 93ff

2) Allerdings liegen keine klaren Angaben darüber vor, wie unterschieden werden soll, ob die Übereinstimmung zwischen den Ebenen II und III auf Stabilität bzw. Rigidität (diese Interpretation wird ebenfalls angeboten) verweist oder stattdessen auf starke Verteidigungsmechanismen.

heiten : sie können beispielsweise ebenso für 'tatsächliche Aggressivität', wie für 'unbewußte Aggressionswünsche' stehen. [1)]

ad level IV :

"It comprises those interpersonal themes which the patient consistently, significantly, and specifically omits in the three other levels." (S.192)

Da diese Ebene jeder Äußerungsform entbehrt, bleibt hier nur eine indirekte Schlußfolgerung möglich, nämlich über den 'avoidance score'. Das Fehlen bestimmter Themen als solches (omission) läßt sich bei einem fixierten und für alle Ebenen gleichen Variablen-System ohne weiteres feststellen. Ob aber die in den Ebenen I bis III ausgelassenen Variablen tatsächlich 'vermieden' wurden und demnach zwar 'verborgen', tatsächlich jedoch 'vorhanden' sind, bedarf zusätzlicher Information, wofür Leary 2 Möglichkeiten ins Auge faßt :

---

1) Ein solches Ebenen-Konzept macht also die widersprüchlichen Validierungsergebnisse (vgl.S.37) nunmehr von einer anderen Seite plausibel : nicht das Testverfahren als solches wird dahingehend verstanden, daß es etwas anderes messe als das Kriteriumsverfahren; sondern die durch den Test 'erreichte' (operational definierte) Persönlichkeitsebene steht in einem jeweils verschiedenen 'Ebenen-Kontext' und so auch in je verschiedener Relation zum verwendeten Kriterium. Es wird also hier die gleiche Voraussetzung gemacht, wie weiter oben diskutiert (S. 39), daß nämlich ein projektiver Test eine bestimmte Ebene hervorbringe. Doch wird diese 'Ebene' nicht als solche, sondern erst im Zusammenhang mit den übrigen Ebenen interpretiert.

1. der Vergleich der Antworten mit dem statistisch zu sichernden 'thematic pull' etwa der TAT-Karten (vgl. dazu Fußnote 1), S.46),

2. eine zusätzliche Analyse von selektivem Vergessen und perceptual defense. [1)]

Es handelt sich also um die 'tiefste' Ebene, welche durch die stärksten Verteidigungsmechanismen ausgezeichnet ist. Sie ist indessen auch von Leary - weil am schwersten zugänglich - bisher nur theoretisch konzipiert und noch nicht praktisch in eine Diagnose mit einbezogen worden.

ad level V :

Diese Ebene wird definiert durch die "ideals bild by the individual - his conceptions of 'rightness', 'goodness', of what he should like to be" (S.200).

Entsprechend der Ebene II handelt es sich hier nur um das 'öffentliche' Ideal- oder Wertsystem, dem das private Ich-Ideal durchaus entgegenstehen kann. Auch diese Ebene gewinnt eine fundamentale Rolle erst im Vergleich mit den übrigen Ebenen. Die Relation insbesondere zwischen den Ebenen II und V wird als ein Index der self-acceptance verstanden. Zugang zu dieser Ebene liefern u.a. Interview und check list.

4. Vergleich des Ansatzes von LEARY mit dem der übrigen Autoren.

Für Leary wie für Rosenzweig ist das Konzept der Ebenen nicht etwa nur ein Beitrag zu einer noch unzulänglichen Test-

---

1) Eine solche Verwendung von perceptual defense wird jedoch nur als eine zukünftige Möglichkeit angedeutet und die diagnostische Valenz von selektivem Vergessen nur anhand einer (damals unveröffentlichten) Arbeit erläutert : "The serial order of recall is believed to relate to intrapsychic defensive processes holding sensitive themes from awareness" (S. 198).

theorie; sondern das Ebenen-Konzept wird hier als ein unerläßliches Mittel für die diagnostische Beurteilung einer Person erachtet. Nicht die Tests verlangen eine begriffliche Ordnung in der Terminologie von 'Ebenen'; sondern die Persönlichkeitsorganisation selbst wird erst verständlich, wenn man sie vermittels eines Ebenen-Modells begreift. Der Begriff der Ebene dient also nicht dazu, verschiedenartige Testergebnisse 'plausibel' zu machen; er präzisiert vielmehr die Aufgabe des Diagnostikers dahin, daß das Verhalten einer Person nicht anders als anhand systematisch nach 'Ebenen' variierter Information beurteilt werden kann.

Das bei Rosenzweig noch vorläufig formulierte Programm, die Diagnose einer Person am Gefüge ihrer Ebenen auszurichten, wird bei Leary zu einer ausgefeilten Theorie. [1] Der bei Rosenzweig noch 'statische' Begriff der Ebenen-Organisation erfährt bei Leary vor allem durch den dynamischen Begriff der 'variability' eine entscheidende Ergänzung. Die Ebenen selbst, wie ihr gegenseitiges Gefüge werden hier vermittels operationaler Definitionen metrisch faßbar; der Grad ihrer Übereinstimmung oder Differenz (definiert als der Grad der 'variability') kann nun zum Ausgangspunkt von Erwartungen und diagnostischer Folgerungen gemacht werden. [2] Auch Begriffe wie Rigidität, Konflikt oder Projektion erhalten ein operationales Korrelat durch die quantitativen Beziehungen zwischen den Ebenen.

Die Theorie von Leary erfüllt also einmal - wie vorstehend deutlich wurde - die Forderung, einzelne Ebenen präzise zu formulieren und zugleich die weitere Aufgabe, die diagnosti-

---

1) Leary scheint jedoch die Arbeit von Rosenzweig nicht zu kennen.

2) Es sei hier nicht im einzelnen auf die Art der Quantifizierung auf den einzelnen Ebenen eingegangen und ebenso nicht auf die Fülle der Variabilitäts-Indizes, welche dadurch entsteht, daß der Proband nicht nur sich selbst, sondern auch Eltern, Therapeuten und andere anhand der vorgegebenen Eigenschaften einschätzt.

sche Relevanz dieser Ebenen genau anzugeben. Zu fragen aber ist nun, welchen Beitrag diese Theorie zum Problem der Funktionsweise projektiver Tests liefert. Hier zeigt sich nun die Schwäche dieses Ansatzes. Der ausdrückliche Anspruch, die Persönlichkeitstheorie nicht auf postulierten 'Kräften' und 'Mechanismen' zu gründen, sondern auf streng operational definierten Begriffen, führt zu einer unkritischen Bewertung dieser Operationen selbst. Ebene III, u.a. definiert als die Ebene des 'imagined self in the preconscious or symbolic world', soll zwar einerseits das imaginative oder symbolhafte Verhalten bezeichnen, wie es sich auch in Fantasie und Traum niederschlägt. Diese Ebene wird praktisch aber operational gleichgesetzt mit der Reaktion auf projektive Tests : "Level III behavior should ... be accurately labeled as the 'response to projective stimuli' ". (S.155)

So wird denn auch hier wie bei den anderen Ebenen-Theoretikern für die projektiven Tests eine bestimmte Funktionsweise unkritisch vorausgesetzt; es wird ein bestimmter Testtyp einer Persönlichkeitsebene zugeordnet, die Ebene durch den Test definiert. [1] Auf der einen Seite wird Ebene III zwar nominal definiert durch Begriffe wie 'symbolic world' etc. Die operationale Definition dieser Begriffe aber auf der anderen Seite bringt keinerlei inhaltliche Klärung des Gemeinten : denn die Prozesse oder Mechanismen, welche durch die projektiven stimuli ausgelöst werden, bleiben ihrerseits ungenannt. Erst die Kenntnis dieser R e a k t i o n e n jedoch könnte eine Grundlage für die Annahme liefern, daß ein projektiver Test durchgängig eine bestimmte Ebene repräsentierte (vgl. dazu S. 40f).

1) Es wird zwar eingeschränkt, daß Verteidigungsverhalten den Ausdruck privater Inhalte hindern könne. Das führt jedoch nicht zu einer Relativierung der Aussagemöglichkeiten proj. Tests; es wird vielmehr festgelegt, daß diese Ebene in solchem Falle besonders 'tief' liege. Ein proj. Test mißt daher ex definitione eine bestimmte 'Tiefe'; die 'privaten Inhalte' liegen jedoch u.U. noch unterhalb dieser.

Mit Rosenzweig ist stattdessen deutlich geworden, daß das Verhalten auf projektive Tests und selbst auf einen einzelnen bestimmten Test mehrdeutig ist. Die operationale Gleichsetzung von der dem Testverhalten als zugrundeliegend gedachten 'psychischen Realität' und einer bestimmten Persönlichkeitsebene macht somit das Definiens unspezifisch; der bezeichnende Begriff geht weit über den bezeichneten Gegenstand hinaus. Leary könnte allenfalls sagen : wenn ein projektiver Test auf der Ebene III funktioniert, dann bezeichne ich dieses Verhalten 'als auf Ebene III sich abspielend.' Dies jedoch ist fraglos zirkulär. Solange die Operationen (als Definiens) nicht eindeutig bestimmbar sind, können die Begriffe, also die 'Ebenen', nur unabhängig von ihnen, hier also unabhängig vom jeweiligen Testmedium, definiert werden. Erst dann ist zu prüfen, welche Operation dem theoretischen Begriff am nächsten kommt, erst dann ist zu entscheiden, welche Funktionsweise einem bestimmten Test zuzuordnen ist. Dieser Punkt wird weiter unten eingehender diskutiert werden (vgl.S. 99ff).

Leary geht allerdings in einem wichtigen Punkt über die sonst üblich gewordene Einstufung projektiver Tests auf einem imaginären 'Tiefenkontinuum' hinaus. Er bezieht nämlich inter-individuelle Unterschiede mit ein, die nicht den m a n i f e s t e n Inhalt der Testantworten betreffen, sondern welche durch das je besondere Verhältnis dieser 'Ebene III' zu den übrigen Ebenen bedingt sind.Auf diese Weise kann aufgedeckt werden, daß im besonderen Falle auf dieser Ebene Inhalte in 'symbolischer' Form auftreten, die sich auf keiner der anderen, also den stärker bewußtseins-kontrollierten Ebenen, finden; oder aber,daß bestimmte Inhalte selbst auf dieser Ebene vermieden wurden, was auf starke Verteidigungsmechanismen schließen läßt. Ein in diesem Sinne differenzierteres Verständnis der projektiven Testantworten ersetzt jedoch nicht die solcher Interpretation vorgeordnete Frage, o b der Test tatsächlich 'indirektes' Verhalten hervorgerufen hat. Nimmt man beispielsweise in einem fiktiven Falle an, daß die Reaktionen auf Ebene III denen auf Ebene I entsprechen, so wäre es denkbar, daß hier weder besondere

'private' Inhalte fehlen, noch daß besonders starke Verteidigungsmechanismen sie verdeckten, sondern daß der Test die 'Tiefendimension' nicht erreichte, welche aber als solche durchaus das Individuum charakterisiert.

Die bisherige Ebenen-Diskussion läßt sich nun dahingehend zusammenfassen, daß der Ansatz von Leary ebensowenig wie das vage Konzept eines 'Tiefenkontinuums' zu e r k l ä r e n vermögen, weshalb verschiedene projektive Tests derart verschiedene Ergebnisse hervorbringen, daß sie - ohne eine eingehendere Analyse - sich nicht miteinander vereinbaren lassen. Rosenzweig allein schließt in seine Konzeption des Ebenen-Begriffs die Erwartung mit ein, daß bereits e i n Test sehr verschiedenartige R e a k t i o n e n hervorbringen könne, daß in einem Test - aller einfachen Zuordnung entgegen - verschiedene Persönlichkeitsbereiche zum Ausdruck gelangen können. Die Ebenen lassen sich hier also als 'Reaktionsebenen' verstehen, für welche der Test als Auslöser fungiert. [1)]

Demgegenüber ist der Ansatz Leary's für eine Testtheorie unfruchtbar. Denn Leary systematisiert nicht 'Verhalten' mit dem Ebenen-Begriff; sondern er faßt die Ebenen als 'Bereiche' auf und reduziert zugleich deren Definition operational auf bestimmte Erhebungstechniken. Der Begriff der 'Ebene' bezeichnet hier nicht verschiedene Formen der 'Reaktion', sondern verschiedene 'Quellen' der 'Information'. Der projektive Test wird bei Leary zum Definiens, so daß er zum Gegenstand der Untersuchung nicht mehr gemacht zu werden braucht.

---

1) Rosenzweig vernachlässigt allerdings die Frage, ob nicht die besondere stimulus-Struktur und stimulus-Relevanz (für diese Unterscheidung der formalen und inhaltlichen Seite des stimulus-Materials vgl. Murstein 1959) eines Tests dazu führen könnte, daß dieser Test vorrangig das Verhalten nur einer 'Ebene' hervorruft. So weisen beispielsweise verschiedene Autoren nach, daß der SET (der ja als 'projektives' Verfahren angesehen wird) eine zu starke Bewußtseinskontrolle provoziere, als daß er latente Inhalte überhaupt hervorbringen könnte (vgl. Kass 1956, Holsopple, Miale 1954; Fitzgerald 1958; Stone, Dellis 1960, Theiner 1962).

Eine Analyse der Funktionsweise projektiver Tests jedoch ist die Aufgabe dieser Arbeit. Es sind daher nun jene Untersuchungen heranzuziehen, welche - von anderen theoretischen Voraussetzungen ausgehend - die Reaktionen auf projektive stimuli direkt zum Gegenstand haben. Erst die bisher e x p e r i m e n t e l l erarbeiteten Kenntnisse über die Funktionsweise projektiver Verfahren, nämlich eingehende Analysen zum Verhältnis von Index und Indiziertem bei diesen Tests, bieten eine Grundlage für eine Testtheorie. Sie bieten u.U. - eben dieses wird zu prüfen sein - einen Ansatzpunkt dafür, in Anlehnung an Rosenzweig das Konzept der 'Reaktionsebenen' soweit theoretisch wie experimentell zu begründen, daß sich schließlich Hypothesen herleiten lassen, welche ihrerseits zur Planung eines Experiments führen können.

II. Forschungsansätze zum Problem des Verhältnisses von Index und Indiziertem soweit sie relevant sind für das Ebenen-Problem.

1. Vorbemerkungen

Für den Zugang zum Problem des Verhältnisses von Index und Indiziertem werden methodisch im wesentlichen 3 Richtungen eingeschlagen : es wird einmal untersucht, in welcher Weise experimentell induzierte Motive sich auf die Testreaktion auswirken; zum anderen werden die Motiv-Indizes eines projektiven Tests verglichen mit (soziometrischen) Einschätzungen des overt-behavior (welches also hier als 'Außenkriterium' fungiert); schließlich wird eine einzelne Motivvariable untersucht, indem man die Motiv-Indizes verschiedener Erhebungstechniken miteinander vergleicht.

Theoretisch werden dabei unterschiedliche Modelle zur Beschreibung und Erklärung der Relation von Motiv und Motivausdruck einerseits, von Testverhalten und offenem Verhalten andererseits herausgebildet : die an Rotter's 'social-learning-theory' anknüpfende Erwartungstheorie, das an Miller anknüpfende 'approach-avoidance-conflict-Modell', die an der social-perception-Forschung anknüpfende Annahme von 'defense und sensitization'. Unterscheiden lassen sich diese Arbeiten weiterhin auch danach, ob sie ansetzen bei der Analyse der Funktionsweise projektiver Tests selbst oder ob der Schwerpunkt - in einer eher allgemeinpsychologischen Orientierung - auf der Suche nach einem methodischen Zugang zur menschlichen Motivation liegt.

Die E b e n e n - Theoretiker nun (mit Ausnahme von Rosenzweig) gehen - so war oben dargelegt worden - davon aus, daß projektive Tests die 'latenten Inhalte' einer Person, ihre 'tieferen' Bereiche zu erschließen vermögen. Die Test-Indizes, die mit einem projektiven Verfahren gewonnen werden, müssen sich demnach systematisch von den Indizes des beobachtbaren Verhaltens und von denen der 'publizierten' Meinungen und Einstellungen (vom 'öffentlichen Selbst') unterscheiden. Die experimentellen Untersuchungen sind also

daraufhin zu prüfen, wieweit sie diese Annahme bestätigen, begründen und u.U. differenzieren, indem sie den Ausdruck latenter (und anderer) Inhalte in seinen Bedingungen analysieren.

2. Allgemeine Relationen zwischen Testindex, Motiv und offenem Verhalten

Eine besondere Rolle in der Analyse der Motivation und ihrer Messung spielen die Arbeiten von Atkinson, McClelland und Mitarbeitern [1]. Bei ihrer umfassenden Untersuchung des achievement-Motivs kommen sie zu dem Ergebnis, daß für eine Motivmessung dem Fantasieverhalten eindeutig der Vorrang gebührt gegenüber jedem anderen Zugang zur Motivation. Der enge methodische Ansatz einer experimentellen Bedürfnisinduktion und -variation, deren Niederschlag dann im projektiven Test sehr sorgfältig gemessen wird, führt jedoch zu einem sehr einfachen Grundprinzip : dem der 'direct expression' [2]. Es wird ein direktes Verhältnis zwischen induzierter Motivstärke und (projektivem) Fantasie-Index nachgewiesen. Für das Beispiel der Leistungsmotivation wäre also hier aufgezeigt, daß ein projektiver Test die aktuelle Motivation [3] widerspiegelt und Vpn ihrer Motivationsstärke nach inter-individuell differenziert.

Die Vorgehensweise einer experimentellen Bedürfnisinduktion und die Allgemeingültigkeit der Ergebnisse ist jedoch von anderen Autoren überzeugend kritisiert worden. Dem Prinzip einer 'direct-expression' stellen sie ein 'alternative-channel-modell (Broverman 1960) bzw. eine 'substitutive-defense-conception' (Lazarus 1961) gegenüber. Broverman weist experimentell ein umgekehrtes Verhältnis von Motivindex und einem Außenkriterium nach, welches hier nicht in

---

1) McClelland et.al.1953; McClelland 1958; Atkinson 1958; Atkinson, Litwin 1960; Atkinson 1961.

2) Es ist hier nicht der Ort, die umfassende Verhaltens- und Motivationstheorie dieser Autoren näher auszuführen. Auch auf eine spätere theoretische Ausweitung dieses Prinzips wird erst im Verlaufe der Argumentation verwiesen werden.

3) vgl. die begriffliche Unterscheidung von 'motive' als 'stable disposition' und 'motivation' als 'temporarily aroused state' bei Atkinson 1958.

einer Laboratoriumssituation (wie bei McClelland et.al.), sondern aus Lebensdaten gewonnen wird. Dieses Ergebnis wird in seinem Unterschied zu demjenigen bei McClelland et.al. von beiden Autoren folgendermaßen interpretiert [1]. Ein Bedürfnis gelangt gerade nur dann in Fantasie-Antworten zum Ausdruck, wenn es im alltäglichen Leben keine Befriedigung erfährt und aus diesem Grunde auf alternative Ausdruckskanäle 'angewiesen' ist. Findet andererseits ein Bedürfnis in der sozialen Umwelt adäquate Bedingungen für seine Realisation, so taucht es nicht noch einmal im Fantasie-Verhalten auf, da die lebensferne Testsituation für es irrelevant ist. Durch die gleiche Irrelevanz aber, so folgern die Autoren, ist auch die experimentelle Situation künstlicher Bedürfnisinduzierung bestimmt. Denn auf grund der relativen Einfachheit der hier verwendeten Aufgaben ist sie nicht weniger weit von der tatsächlichen Lebens- (und Problem-) Situation entfernt wie die Testsituation selber. Ein im sozialen Leben nicht realisiertes Leistungsstreben kann in der experimentellen Kriteriums-Situation (Aufgaben-lösen) genau wie in der Testsituation (TAT) einen 'Ausdruckskanal' finden, so daß die Testreaktionen tatsächlich dem Verhalten im Experiment (Kriterium), beide jedoch nicht der Lebenssituation entsprechen, sondern im umgekehrten Verhältnis zu ihr stehen. So entspricht dem substitutiven Ausdruck nichtrealisierter Bedürfnisse im Fantasieverhalten ein ebenso substitutives Leistungsstreben in den - gegenüber den tatsächlichen Anforderungen der sozialen Umwelt - als relativ einfach empfundenen Aufgaben. Diese dienen demnach in ganz unangemessener Weise als Außenkriterium.

---

1) Die berechtigte Kritik an dieser Arbeit (Atkinson 1961, Heckhausen 1963), daß nämlich für das hier verwendete Fantasiemaß keine Reliabilitäts- und Validitätsangaben vorliegen, muß in diesem Zusammenhang nicht mit berücksichtigt werden, da es hier mehr um die grundsätzliche Diskussion von 'Modellen' geht. Das Modell selbst wird nämlich durch diese Kritik nicht tangiert, sondern nur die experimentelle Realisation der aus ihm hergeleiteten Annahmen.

Diesem alternativ-channel-modell zur Folge sind es also nicht Motive schlechthin, welche in den Reaktionen auf projektive stimuli zum Ausdruck gelangen; vielmehr sind es vorrangig die unbefriedigten Bedürfnisse. Erst die Differenz zwischen Bedürfnis und seiner Realisation oder die zwischen Leistungsanspruch und tatsächlichem Verhalten führt somit zum Ausdruck des betreffenden Motivs, während - in den Begriffen des Ebenen-Konzeptes gesprochen - die Übereinstimmung zwischen den 'Ebenen' einen solchen Ausdruck nicht zur Folge hat. Nicht die Stärke des Motivs für sich genommen ist ausschlaggebend, sondern die Motivstärke im Verhältnis zur verhaltensmäßigen Realisation dieses Bedürfnisses, oder, in den Termini des Ebenen-Konzeptes : das Motiv, bezogen auf das G e f ü g e [1] der übrigen Ebenen. [2]

Es zeigt diese erste Gegenüberstellung kontroverser Modelle also schon so viel an, daß das Verhältnis von Index und Indiziertem bei projektiven Tests nicht durch einfache Zuordnungen hinreichend zu bestimmen ist. Weitere Differenzierungen, in seiner begrifflichen Beschreibung, wie in der experimentellen Methodik, erweisen sich als notwendig.

Welches Ausmaß an differenzierter Bedingungsanalyse diese Fragestellung notwendig macht, sei an einem weiteren Beispiel aus den Arbeiten zur Leistungsmotivation vorgeführt. In einer Untersuchung von Lesser, Krawitz und Packard (1963) führt die experimentelle Induktion von achievement-Motivation nicht dazu, daß die Motiv-Indizes in Fantasieproduktionen (TAT-ähnliche Bilder) gegenüber jenen in einer neutralen Situation

---

1) Der Begriff des Ebenen-Gefüges verweist fortan darauf, daß gleichzeitig alle Ebenen in ihrer Relation zueinander betrachtet werden. Leary (1957) verwendet den analogen Begriff des 'multilevel pattern'; Rosenzweig (1950, S.69) spricht von 'integration of ... personality in terms of ... levels'.

2) Diese Auffassung einer substitutiven Rolle der Fantasie wird bei Lazarus jedoch ergänzt durch die Annahme, daß insbesondere geringe Motiv-Indizes nicht nur auf ein 'befriedigtes', sondern auch auf ein gering ausgeprägtes Motiv, schließlich aber auch auf Verteidigungsverhalten verweisen können.

erhöht sind, wie nach dem direct-expression-Modell zu erwarten wäre. Sondern diese Erwartung wird nur für männliche, nicht für weibliche Vpn bestätigt. Zwar spiegelt das Fantasiemaß bei den weiblichen Vpn denselben Unterschied zwischen 'highachievers' und 'underachievers' wider, wie er an einem Außenkriterium festgestellt wurde; doch zeigen beide Gruppen (die weibl. high- und underachievers) nicht den erwarteten Anstieg der achievement-scores im Vergleich von neutraler und achievement-gerichteter Situation. Nur durch eine zusätzliche Differenzierung kann die Erwartung bestätigt werden. Es zeigt sich nämlich, daß man bei dieser Gruppe weiblicher Vpn die verwendeten TAT-ähnlichen Bilder danach unterscheiden muß, daß auf den einen männliche Figuren, auf den anderen weibliche Figuren abgebildet sind. Die highachievers zeigen den erwarteten Anstieg der achievement-Fantasie-Indizes nur bei den Bildern mit weiblichen Figuren, die underachievers nur bei den Bildern mit männlichen Figuren : in beiden Fällen ein Anstieg, der bei fehlender Differenzierung nach Art der Bilder ausgeglichen und nicht sichtbar wird. Eine plausible Erklärung findet dieses Ergebnis erst durch die zusätzliche Variable der 'Rollenerwartung', der Rollenerwartung nämlich i.b. auf den 'Helden' : Weibliche highachievers sehen ihre durch Orientierung am akademischen Erfolg bestimmte Rolle repräsentiert durch weibliche Gestalten im stimulus-Material, so daß durch diese der Ausdruck erhöhter (induzierter) Motivation ausgelöst wird; weibliche underachievers dagegen verbinden (bei gleichem Intelligenzgrad) Leistungsstreben mit der Rolle des Mannes, weshalb bei ihnen männliche 'Helden' zu Auslöserreizen für den Ausdruck leistungsorientierter Motivation werden.

Diese Untersuchung ist in mehrfacher Hinsicht aufschlußreich. Es ergibt sich nämlich, daß dieselben stimuli von sehr unterschiedlicher Relevanz sein können und daß diese ihre unterschiedliche Rolle bestimmt wird gerade durch die Eigenart der Motivstruktur der Vp. Unterschiedliche Motivstärke ist hier mit unterschiedlichen kognitiven Erwartungen verknüpft, welche ihrerseits zu differenten Reaktionen auf das stimulus-Material führen. Kognitive Einstellungen entscheiden also über die Art des Motivausdrucks : nur die-

jenigen stimuli, die diesen Einstellungen korrespondieren, lassen das Motiv selbst überhaupt erkennen. Bei den underachievers jedoch, bei Vpn also mit besonders geringer Leistungsmotivation, sind es die kognitiven Einstellungen a l l e i n (da eine Motivation kaum besteht), welche bei entsprechenden stimuli, nämlich bei den Einstellungen entsprechenden stimuli (männliche 'Helden'), leistungsbezogene Antworten hervorbringen. Diese Rolle kognitiver Einstellungen bei Fantasie-Reaktionen betonen auch Moss und Kagan (1961) : "...that the production of thematic material is strongly influenced by the subject's conception of what behaviors are appropriate to the hero's social role" (S.63). Es sind demnach also auch 'Oberflächenerscheinungen', welche die Reaktionen auf projektive stimuli ausschließlich bedingen oder doch entscheidend modifizieren können.

Bei einigen anderen Autoren wird der Begriff der Erwartung in einem weiteren Sinne verwendet ; er dient hier als allgemeines Erklärungsprinzip für das Verhältnis von Index und Indiziertem. Dieser Begriff der 'expectancy' ist an der social learning theory von Rotter (1954, 1955) orientiert : durch das stimulus-Material, wie durch die Test-Situation werden unterschiedliche 'Erwartungen' darüber hervorgerufen, welcherart Bedürfnisbefriedigung das geforderte Testverhalten verspricht. Die Reaktion wird schließlich bestimmt durch das jeweilige Zusammenspiel möglicher (antizipierter) Bedürfnisbefriedigungen und der objektiven Aussicht auf tatsächliche Befriedigung. Nach Fitzgerald (1958) legen projektive Tests vor allem bei nichtrealisierten Bedürfnissen eine 'Erwartung' auf Befriedigung nahe. Dieses Erklärungsprinzip ist allerdings heute noch kaum weniger allgemein und vage als das verschieden starker Bewußtseinskontrolle (welches zur Vorstellung eines Tiefenkontinuums führte). Auch mit ihm kann 'begründet' werden, daß verschiedene Tests unterschiedliche Ergebnisse erbringen müssen, weil sie ja unterschiedliche 'Erwartungen' hervorrufen (Marlowe 1959).

Einen ähnlichen Begriff von 'expectancy' verwendet Atkinson (1958) in Anlehnung an den Begriff des goal attainment bei

Tolman. Er dient hier dazu, den Abstand zwischen dispositionellem Motiv und aktueller Motivation (vgl.Fußnote 3),S.55) zu überbrücken : die cues von Testmaterial und Situation sprechen die latenten dispositionellen Motive an;, es entstehen aktuelle Erwartungen über eine Bedürfnisbefriedigung (goal attainment) und somit ein Zustand aktueller Motivation, so daß von dessen Ausdruck im Fantasieverhalten zurückgeschlossen werden kann auf das dispositionelle Motiv selbst.

Einen noch differenzierteren Beitrag zu der Frage, weshalb überhaupt motivationale Inhalte in den Reaktionen auf projektive stimuli zum Ausdruck gelangen, liefert Fuchs (1958). Nach ihm spricht das Motiv auf die entsprechenden Hinweisreize schon vor deren kognitiver Erkennung an : die motivbezogenen Erwartungsschemata lösen nämlich eine Affektaktivierung aus, welche dann ihrerseits auf die Wahrnehmungsorganisation steuernd, nämlich - im Sinne des Affekts - selegierend einwirkt. Diese Hypothese muß jedoch in ihrem Erklärungswert auf den experimentellen Sonderfall der Motiv-Induzierung beschränkt bleiben [1]. Die Bildung der motivbezogenen Erwartungsschemata geht nämlich hier zeitlich der Darbietung der Teststimuli voraus; sie sind bereits (in den durchgeführten Experimenten) a k t u e l l vorhanden. Die Frage, über welchen Mechanismus in einer demgegenüber neutralen Situation (also ohne vorhergehende Motiv-Induzierung, wie z.B. bei einer diagnostischen Situation) bestimmte Motive durch die Test-stimuli angesprochen werden und dann die Testreaktion bestimmen, wird auch hier nicht (jedenfalls nicht experimentell) geklärt.

Durch den Begriff der 'expectancy' kann zwar umschrieben werden, d a ß Motivation, Wahrnehmung und Reaktion eng miteinander verknüpft sind. Es wird 'plausibel', daß unterschiedliche Tests verschiedenartige 'Erwartungen' hervorrufen können, daß Vpn sich hinsichtlich ihrer Erwartungsstruk-

1) abgesehen einmal davon, daß die Annahme selbst kritisch zu analysieren wäre, nach welcher die Motivaktivierung der Erkennung des Aktivierungsreizes vorangeht.

tur inter-individuell unterscheiden. Auch bekräftigt die expectancy-Hypothese die Annahme, daß Tests tatsächlich (indem sie ein Mittel der Bedürfnisbefriedigung sind) Motivationales zum Ausdruck bringen und so dem diagnostischen Zugriff zugänglich machen. So gesehen, liefert der Begriff einen Beitrag dazu, den Vorgang der Testreaktion zu beschreiben. Dieser Ansatz führt jedoch nicht zu Hypothesen darüber, welcherart Motive zum Ausdruck gelangen, unter welchen Bedingungen unbefriedigte, konfliktbesetzte oder befriedigte Motive sich in den Testreaktionen widerspiegeln, wann und welcherart Motive etwa wirksam vermieden werden oder wann etwa nur kognitive Einstellungen in den Testantworten wiedergegeben werden.

Die bisher beispielhaft zitierten Untersuchungen weisen daraufhin, daß zweierlei näher untersucht werden muß, soll das Zustandekommen der Test-Reaktion in seinen Bedingungen erfaßt werden : die Rolle des Test-stimulus als Auslöser der Reaktion und das Motiv im weiteren Kontext von Verhalten und Verhaltensanspruch (z.B. Leistungsanspruch). Es sind also im folgenden Arbeiten heranzuziehen, welche die Genese der Test-Reaktion in derart differenzierterer Form analysieren.

3. Zur Genese der Test-Reaktion

a) Zur Rolle von approach- und avoidance-Tendenzen

Vermochte der Begriff der 'expectancy' das Verhältnis von dispositionellem Motiv und seinem aktuellen Ausdruck, zwischen stimulus und seiner Wahrnehmung nur allgemein und ohne Hinweis auf bestimmte zu erwartende experimentelle Relationen zu umschreiben, so ist es mit der Einführung des an Miller (1944) orientierten approach-avoidance-Konflikt-Modells möglich geworden, diese Zusammenhänge theoretisch so zu formulieren, daß empirisch prüfbare Hypothesen ableitbar wurden. [1)]

---

1) vgl. Goss, Brownell 1956; Epstein, Smith 1956; Scott 1956; Weisskopf-Joelsen et.al. 1957; Lesser 1958,1959b; Atkinson, Litwin 1960; Phares 1961.

Insbesondere das Verhältnis zwischen den lebensfernen Teststimuli zu jenen Reizgegebenheiten, welche im alltäglichen Leben mit der Befriedigung oder Versagung eines Bedürfnisses verknüpft sind (extra-Test-stimuli), wird nun hier durch den Begriff der 'Reizgeneralisation' in einem präzisen theoretischen Zusammenhang formuliert (Goss und Brownell 1956).

Anhand dieses Begriffs läßt sich - dem Modell nach - beschreiben, in welchem Maße die Test-stimuli die Reaktion determinieren. Der Grad der Übereinstimmung zwischen Test-stimulus und extra-Test-stimulus (der Grad der 'stimulus-relevance' [1]) ist ausschlaggebend dafür, mit welchem Grad an Motiv-Ausdruck - bei gegebener Motivstärke - zu rechnen ist (siehe dazu weiter unten).

Mit den Begriffen 'approach' und 'avoidance' werden jene Tendenzen beschrieben, deren Widerstreit zum 'Konflikt' führt : einerseits der Wunsch, ein Bedürfnis zu befriedigen (Hinwendungsreaktion) und andererseits der Wunsch (oder auch der Anspruch), diese Befriedigung zu verhindern (Abwehrreaktion). In der Lebens- wie in der Testsituation entscheidet sich - so beschreibt es das Modell - die Reaktion danach, welche der beiden Tendenzen überwiegt.

In der Testsituation hängt die Stärke beider Tendenzen einmal von der Bedürfnisstärke selbst ab und zugleich vom Grad der stimulus-relevance (s.o.). Das Zusammenspiel beider Tendenzen ist dadurch bestimmt, daß der avoidance-Gradient prin-

---

1) Der Begriff der 'stimulus-relevance' ist als analog zu verstehen zu dem Begriff der 'Entfernung vom Ziel' in den Experimenten zum tierischen Vermeidenslernen, aus welchen das Konfliktmodell hervorging. Bei den meisten Autoren ist jedoch die Anlehnung an dieses S-R-theoretische Modell nicht sehr streng : es werden daher Begriffe wie 'cue strength', 'content pull', 'ambiguity' dem Begriff der 'stimulus-relevance' synonym verwendet. Auch wir werden zunächst keine klare begriffliche Scheidung vornehmen, da die S-R-theoretischen Implikationen des Modells bei den meisten Autoren unberücksichtigt bleiben. Doch wird die Rolle des Begriffs der 'stimulus-relevance' als S-R-theoretisches Konzept an späterer Stelle kritisch zu analysieren sein (s.S.75f und S.93ff).

zipiell 'steiler' ist als der approach-Gradient; das besagt : mit zunehmender Zielnähe (sprich 'Aussicht auf Befriedigung'), also mit zunehmender stimulus-relevance nimmt die avoidance-Tendenz stärker zu als die approach-Tendenz. Angewendet auf ein Testverfahren umfaßt das Konfliktmodell demnach folgende Variablen : stimulus-relevance, Motivstärke, approach- und avoidance-Tendenz, Motivausdruck (repräsentiert durch den Motiv-Index). Der funktionale Zusammenhang dieser Variablen liefert - dem Modell nach - eine vollständige Analyse der Relation von Reiz (Test-stimulus) und Reaktion (Testverhalten). [1)]

Heuristisch ist die Orientierung am Konflikt-Modell insbesondere deshalb fruchtbar, weil hier erstmalig der Gesichtspunkt der V e r m e i d u n g von Motivausdruck systematisch in die Analyse der Testantworten einbezogen wird. Bei der Annahme allein von Projektionsvorgängen konnte er nicht in den Blick geraten (s.S. 18f). Aber auch bei den Untersuchungen zur Leistungsmotivation (s.S. 55ff) brauchte er nicht in Betracht gezogen zu werden, da (defensive) 'Vermeidung' bei einem sozial so positiv bewerteten Motiv nicht von Belang ist.

Schon 1950 hatte Pitluck darauf aufmerksam gemacht, daß das Verhältnis zwischen Fantasie-Index für Aggression und offener Aggression sich danach entscheidet, ob Aggressionsangst und somit Aggressionsvermeidung vorliegen oder nicht. Nicht die Frage, ob ein Bedürfnis befriedigt ist oder nicht, sondern auch die, ob es 'sozial zugelassen' bzw. angst- oder schuldbesetzt ist oder nicht, wird somit zu einem entscheidenden Kriterium. Diese Vermeidung der 'tieferen' (latenten) Inhalte selbst bei einem Test wie dem TAT wurde von den Ebenen-Theoretikern nicht systematisch mit einbezogen; nur bei Leary werden die derart durchgehend vermiedenen Inhalte als eine eigene Ebene (IV) konzipiert. (vgl.S. 47f)

---

1) Auf weitere Differenzierungen, wie sie z.B. bei Goss und Brownell (1956) eingeführt wurden (etwa die Unterscheidung zwischen primary stimulus generalization und response mediated generalization) sei aus den in Fußnote 1),S.62 genannten Gründen nicht eingegangen.

Theoretisch läßt sich für den Fall der Konfliktbesetzung, also des Widerstreits zwischen approach- und avoidance-Tendenz bei e i n e m Motiv, aus diesem Modell ableiten, daß je eindeutiger die cues, desto stärker auch die Vermeidensreaktionen sind und umgekehrt. Die Stärke der Vermeidensreaktion variiert (ebenso wie die Hinwendungsreaktion) zugleich mit der Stärke des Motivs selbst; und so folgt, daß, je stärker der Trieb ist, eine um so geringere cue-Stärke notwendig wird, um die defensiven Vermeidensreaktionen auszulösen. Besteht andererseits eine solche Konfliktbesetzung nicht und wird der Test-stimulus auf einen extra-Test-stimulus hin generalisiert, mit dem eine Hinwendungsreaktion, also die Befriedigung des Bedürfnisses, assoziiert ist, so besteht nach diesem Modell bei zunehmender cue-Stärke eine zunehmend positive Relation zwischen Bedürfnis und seinem Ausdruck.

Diese aus dem Modell formal ableitbaren Konsequenzen sind jedoch für seine Anwendung auf das Testverhalten einzuschränken. Es ist nämlich in Zweifel zu ziehen, ob auch e i n d e u t i g e stimuli notwendig Motivationales hervorrufen. Es könnte stattdessen dem Test gegenüber eine Abwehrhaltung aktualisiert werden, welche nur soziale Stereotypen ohne diagnostische Relevanz zuließe, je offensichtlicher die Absichts des Tests zu Tage tritt. Für ein Motiv, das im Verhalten befriedigt werden kann, wäre die Testsituation nach Brovermann und Lazarus ohnehin irrelevant. Ein konfliktbesetztes Motiv würde - dem Konflikt-Modell zufolge - insbesondere bei hoher stimulus-relevance vermieden. Wenn trotzdem solchen eindeutigen Reizgegebenheiten entsprechend, also im Sinne ihres 'content-pull's' reagiert wird, kann solche Reaktion a l l e i n durch die Test-stimuli determiniert sein und enthielte keine Hinweise auf den Pb selbst und seine Motive. Hier wäre mit Kornardt (1957) zu fragen, ob nicht diese Reaktion - jedenfalls für den Diagnostiker - trivial sei.

Diese Einschränkung ist wichtig, soll das Konflikt-Modell mit jenen Untersuchungen in Einklang gebracht werden können, welche eine kurvilineare Beziehung zwischen Ambiguität und

diagnostischer Relevanz nachgewiesen haben. [1]

Nicht nur für die approach-, auch hinsichtlich der avoidance-Tendenz läßt sich dieses Modell nicht unbedingt vollständig auf das Testverhalten übertragen. Der Begriff der cue-Stärke kann auch hier nicht unbesehen als dem der 'Entfernung vom Ziel' in Tierexperimenten analog verstanden werdem. Ist beim tierischen Vermeidenslernen die Abwehr in Zielnähe am stärksten, so führt eine dem 'Ziel' analoge 'Eindeutigkeit' der Test-stimuli u.U. dazu, daß man sich ihrem 'content-pull' nicht mehr entziehen kann, also mit 'approach' reagieren muß.

Dieses Argument widerspricht nicht dem soeben dargelegten, daß bei hoher stimulus-relevance Triviallösungen erwartet werden könnten. Denn beide Tendenzen, approach und avoidance, sind getrennt zu betrachten. Stimuli mit hoher Relevanz für Achievement bedingen u.U., daß der Pb mit sozialen Stereotypen reagiert; stimuli mit hoher Relevanz für 'Angst' können dazu führen, daß der Ausdruck der eigenen 'Ängstlichkeit' nicht länger defensiv umgangen werden kann (vgl.S. 72). Die Rolle einer hohen stimulus-Relevanz wird im Konflikt-Modell sowohl für die approach-, wie für die avoidance-Tendenz irreführend bestimmt. Material mit stimulus-Eindeutigkeit ist für diagnostische Zwecke weitgehend unbrauchbar; doch sind es stimuli von 'relativ hoher Relevanz', welche aufschlußreiche Reaktionen hervorrufen können

Dennoch hat sich das Konflikt-Modell als heuristisch besonders fruchtbar bewährt, indem es zur Grundlage einer Fülle sehr aufschlußreicher Untersuchungen wurde (vgl.Fußnote 1), S. 61). Allerdings hat es, in Verbindung mit entsprechenden Untersuchungen, teilweise die einseitige Schlußfolgerung angeregt, nicht den Motivausdruck bei relativ unstrukturierten stimuli (wie sie für die gebräuchlichen projektiven Tests charakteristisch sind), sondern dessen Vermeidung bei relativ hoher cue-Stärke als allein diagnostisch relevant anzuerkennen. So hat sich etwa nachweisen lassen, daß Gruppen, die sich in ihrer Aggressivität im overt behavior un-

1) vgl.: Bijou, Kenny 1951; Kenny, Bijou 1953; Murstein 1958b.

terscheiden, nicht durch ihre Reaktion auf unstrukturiertes Material differenziert werden, sondern nur durch den Grad ihrer Vermeidung gegenüber stark aggressions-geladenen cues (hohe cue-Stärke oder stimulus-Relevanz) [1]. Doch wird die Konsequenz solcher Ergebnisse im folgenden (und weiter unten) zu relativieren sein.

Eine solche einseitige Betrachtung nur des Vermeidensverhaltens und eine Validierung von Tests mit relativ hoher stimulus-Relevanz allein auf der Basis von ego-defense bzw. avoidance führt dazu, daß eine Interpretation von a p p r o a c h-Verhalten solchen stimuli gegenüber nicht möglich ist. Handelt es sich in dem Falle, daß Vermeidensverhalten nicht auftritt, sondern dem content-pull entsprechend reagiert wird, um eine diagnostisch triviale Reaktion, oder - der Annahme von direct-expression zufolge - um direkten Motivausdruck oder, im Sinne des alternative-channel-Modells, um die 'substitutive' Befriedigung eines unbefriedigten Bedürfnisses, für welches hier, bei hoher stimulus-relevance, eine hohe 'Erwartung auf Bedürfnisbefriedigung' gegeben sein könnte? Das Verhältnis von stimulus-Struktur, Motivstärke und Motivausdruck ist komplizierter, als daß es durch 'Vermeidung' allein hinreichend aufgeschlüsselt werden könnte. Daher ist insbesondere ein Vorschlag von Farberow (1959) zu beachten : projektive Verfahren sollten so konstruiert sein, daß sie die systematische Messung einzelner Motive dadurch ermöglichen, daß sie ein nach der Intensität der cue-Stärke graduell abgestuftes stimulus-Material bieten. Auf diesen Vorschlag wird weiter unten zurückzukommen sein(vgl.S.80ff).

Eine andere Konsequenz der Anwendung des Konfliktmodells und zwar der Anwendung auf eines der gebräuchlichen projektiven Verfahren (anstatt auf eigens konstruiertes stimulus-Material), zeigt Lesser auf (1958, 1959a,b). Er betrachtet nicht die Wirkung der beiden widerstreitenden Tendenzen in ihrem Ergebnis, das sich entweder als Motivausdruck oder als dessen Vermeidung feststellen läßt. Es weist vielmehr anhand TAT-ähnlicher

1) vgl. : Crandall 1955; Epstein, Smith 1956; Kagan 1956,1959.

Bilder nach, daß beide Tendenzen gleichzeitig und unabhängig voneinander bestimmt werden können und erst in ihrem gegenseitigen Verhältnis diagnostisch aufschlußreich sind. Lesser stellt den Indizes für Fantasie-Aggression (FA) diejenigen Reaktionen gegenüber, welche (in der TAT-Geschichte) für eine mehr oder weniger direkte und effektive Verhinderung dieser Aggressionstendenz stehen und somit Aggressions-Angst (AA) repräsentieren. [1] Erst der Quotient (Konfliktscore) zwischen beiden (FA/AA) führt zu einer hoch signifikanten Korrelation zwischen den (derart kombinierten) Fantasie-Indizes und dem offenen Verhalten : je mehr im Testmaterial die Aggressionsangst (AA) die Aggressionsbereitschaft (FA) überwiegt, desto eindeutiger läßt sich erwarten, daß auch im offenen Verhalten Aggression vermieden wird; je stärker Aggressionsbereitschaft im Verhältnis zur Aggressionsangst zum Ausdruck gelangt, um so mehr ist im offenen Verhalten Aggressivität zu erwarten. Ohne diese zusätzliche Variable 'Aggressions-Angst', also ohne Berücksichtigung der avoidance-Tendenz, wären diese Ergebnisse sehr viel weniger aufschlußreich : die Korrelation zwischen offenem Verhalten und Fantasie-Aggression allein ist gering. Leary zieht daraus (1959b) die Konsequenz, widersprüchliche Validitätskoeffizienten weniger den Verfahren zur Last zu legen, als der Verschiedenartigkeit der zugrundegelegten Populationen zuzuschreiben. [2] Hier kommt dem Kriterium 'Vermeidensverhalten' (avoidance) also die sehr präzise Funktion zu, Populationen danach zu differenzieren, ob ein direktes oder indirektes Verhältnis zwischen Fantasieausdruck und offenem Verhalten zu erwarten ist, bzw. eine hohe oder geringe Korrelation zwischen verschiedenartigen Aggressions-Maßen.

---

1) Es werden z.B. die unmittelbare Vereitelung von Aggression, der Abbruch aggressiven Verhaltens und dessen Bestrafung im nachhinein (also innerhalb des Handlungsablaufs der Geschichte) als AA signiert und zusätzlich in der hier angegebenen Reihenfolge gewichtet.

2) Die Rolle derartiger Populations-Differenzen wird später ausführlicher behandelt werden (vgl.S.87ff).

Eine solche Möglichkeit, Vermeidensverhalten im Testmaterial direkt festzustellen, ist auch, wie sich einer Arbeit von Mussen und Naylor (1954) entnehmen läßt, für den TAT selbst gegeben. Diese in den Testantworten sprachlich ausgedrückte Ablehnung von Aggression sei im folgenden als 't h e m a - t i s c h e V e r m e i d u n g' bezeichnet : sie wird für das hier durchgeführte Experiment von Bedeutung sein.

Eine für die Fragestellung dieser Arbeit besonders aufschlußreiche Erweiterung der Anwendungsmöglichkeiten des Konfliktmodells findet sich bei Epstein, der mit seinen Mitarbeitern eine Fülle sehr informativer Untersuchungen durchgeführt hat. [1] Er legt dem Modell nicht jenen Konflikt zwischen approach- und avoidance-Tendenz zugrunde, welcher im alltäglichen Leben für das Motiv charakteristisch ist. [2] Er geht vielmehr aus von dem Widerstreit zwischen der Tendenz, das Motiv bei bestimmten stimulus-Gegebenheiten und in einer bestimmten Situation zum Ausdruck zu bringen (expression) und der gegenläufigen Tendenz, diesen Ausdruck zu vermeiden (inhibition). Bei dieser modifizierten Anwendung erschließt das Konflikt-Modell nicht nur das Fantasieverhalten bei solchen Motiven, die im alltäglichen Leben von 'Schuld' begleitet oder die ich-bedrohlich sind. Sondern es vermag nun auch zutreffender den Ausdruck solcher Motive zu beschreiben, die im alltäglichen Verhalten nicht konfliktbesetzt sind, in der Testsituation aber trotzdem vermieden werden. Es ist dies möglich, weil das Vermeidensverhalten in seinen a k t u e l l e n Bedingungen analysiert wird : nicht nur jene Vermeidung, die im Verhalten für das Motiv charakteristisch ist, wirkt sich hemmend auf den Motivausdruck aus, sondern auch situative Determinanten können ihn verhindern.

Die Vermeidung (hier also 'inhibition') von Motivausdruck hat

---

1) Epstein 1962; Lieman, Epstein 1961; Fenz, Epstein 1962; Saltz, Epstein 1963; Epstein 1966

2) So ging z.B. Lesser vor (vgl.S. 66f) : er stellte die Gruppen mit und ohne Aggressionsangst danach zusammen, ob in ihrer Erziehung Aggression sanktioniert oder zugelassen wurde.

bei Epstein 3 verschiedene Formen : 'reality oriented inhibition', 'drive-reducing inhibition' und 'guilt-reducing inhibition'. Sie seien im folgenden kurz anhand von Beispielen erläutert, welche einem Sammelreferat von Epstein (1962) entnommen sind.

Der 'reality oriented inhibition" liegt nach Epstein "a learned tendency to compensate for drive-related thoughts" zugrunde (1962, S.136). Diese Tendenz, vom eigenen aktuellen Bedürfniszustand gedanklich abzulenken, macht beispielsweise das experimentelle Ergebnis verständlich, daß erst die Daten eines self-report (direkte Befragung!), nicht jedoch projektive Antworten mit dem tatsächlichen Grad an food-deprivation korrelieren. Dieser Zusammenhang verweist zugleich noch einmal darauf, daß nicht eine lineare Relation zwischen stimulus-Relevanz und Vermeidung besteht : bei stimulus-Eindeutigkeit (Befragung) ist Vermeidung u.U. nicht mehr möglich (vgl.S. 65).

Die 'drive-reducing inhibition' bezeichnet eine verstärkte realitäts-orientierte Vermeidung. Die Aussagen werden hier nicht nur einer gelernten Tendenz zufolge umgangen, sondern aus dem zusätzlichen Grunde, daß sie das Motiv selbst verstärken würden. Soll man (in der experimentellen Situation) beispielsweise mit seinem Hunger oder seiner Müdigkeit 'durchhalten', muß man von seinem Bedürfniszustand abzusehen versuchen, soll er sich nicht noch verstärken. Diese Tendenz wird deutlich z.B. darin, daß die Differenz zwischen dem (hohen) Grad an sleep-deprivation und der (geringen) Anzahl Schlaf-bezogener Antworten auf TAT-ähnliche Bilder um so größer wird, wenn zusätzlich bekannt ist, daß dieser Zustand der deprivation noch länger anhalten wird. Epstein folgert daraus, daß "... the TAT reflects the subject's coping reactions to his state of drive, rather than the drive itself." (1962, S. 152) Sowohl aus der expectancy-Theorie, wie aus dem Konflikt-Modell läßt sich jedoch für den Fall einer solchen starken Motivation (deprivation) nur entnehmen, daß der Motivausdruck schon bei geringer stimulus-Relevanz hoch sein müsse. Daß nicht das Motiv selbst, sondern seine 'Kontrolle' die Testreaktion bestimmt, läßt sich mit

diesen Theorien zwar vereinbaren, nicht jedoch aus ihnen ableiten.

Die 'guilt-reducing inhibition' kommt in allen jenen Experimenten zum Ausdruck, in welchen bei 'Schuldgefühlen' oder 'Angst' anstelle eines direkten Verhältnisses von Motivindex und offenem Verhalten ein indirektes Verhältnis festgestellt wird.

Ein Experiment, das sehr deutlich die Notwendigkeit aufzeigt, zwischen 'avoidance' im üblichen Sinne und 'inhibition' im Sinne Epstein's zu unterscheiden und darin besonders aufschlußreich für unsere Fragestellung ist, sei hier noch angeführt, zumal es methodisch beispielhaft differenziert ist. Es werden Motive nicht innerhalb experimenteller, sondern realer Situationen untersucht; es werden 2 projektive Verfahren (Wortassoziation,TAT) mit systematischer stimulus-Variation (3 stimulus-Stärken) [1] dargeboten,und es werden die Motive zu Zeitpunkten verschiedener Stärke gemessen [2]. Das Motiv ist hier 'Angst', gemessen bei Fallschirmspringern in verschiedenen zeitlichen Abständen vom Sprung selbst.

Der üblichen Anwendung des Konflikt-Modells nach kann folgendes erwartet werden : insbesondere bei schwachen cues werden Fallschirmspringer sehr viel mehr Absprung-bezogene Antworten geben als Nicht-Springer (hohe Reizgeneralisation); am Tage des Springens (höchste Angst) werden bei schwachen cues noch Absprung-relevante Antworten auftreten, nicht jedoch bei starken cues (stärkste Vermeidung); 2 Wochen vor dem Absprung aber wird dieses Verhältnis umgekehrt sein. Es zeigt sich jedoch, daß mit zunehmender cue-Stärke eine zu-

1) Dieses Vorgehen erfüllt (zumindest annähernd) die Forderung Holzkamps (1964, S.269), die Ordnungsdimension der 'Stabilität-Labilität' der Wahrnehmungsgegebenheiten vollständig zu repräsentieren.

2) Daß hier ebenfalls auch die Aktivations-Komponente durch den GSR und als Maß für das 'performence deficit' Reaktionszeiten gemessen wurden, bleibe nicht unerwähnt; es ist jedoch für unseren Zusammenhang nicht von unmittelbarem Interesse.

nehmende Relevanz der Antworten verknüpft ist. Dasselbe gilt insbesondere für die Reaktion auf stimuli hoher Relevanz am Tage des Sprungs (also am 'Ziel') : dem Konflikt-Modell nach müßte hier die Vermeidensreaktion am stärksten sein. Indessen tritt kein avoidance-Verhalten auf, nach welchem der content-pull der stimuli defensiv umgangen würde. Vielmehr wird diesem entsprechend reagiert, wobei - und das ist hier entscheidend - nicht die tatsächliche Angst selbst zum Ausdruck kommt (die u.a. durch den GSR gemessen wird), sondern stattdessen Antworten gegeben werden, welche - inhaltlich auf den Fallschirmsprung bezogen - mutiges Verhalten beschreiben. Die tatsächliche Angst wird kompensiert bzw. kontrolliert : ihr Ausdruck würde nur zu ihrer Verstärkung führen. Es handelt sich also um eine drive-reducing inhibition (s.S.69), welche nicht mit der 'avoidance' in der üblichen Verwendung des Konflikt-Modells verwechselt werden darf. Hier nämlich werden nicht die mit den stimuli verknüpften Inhalte als solche umgangen (das Thema des Fallschirmsprungs), sondern der Ausdruck des aktuell vorhandenen 'drive-state's' wird vermieden, indem jene Reaktion beschrieben wird, welche allein situations-angemessen ist (denn es muß ja schließlich gesprungen werden) und welche der Pb demnach von sich erwartet. Nicht die 'tieferen' Schichten kommen zum Ausdruck, sondern 'projiziert' wird der Verhaltensstandard, welcher mit der in den Test-stimuli beschriebenen Situation verknüpft wird.

Von der Voraussetzung ausgehend, daß durch projektive Tests die Inhalte 'tieferer Schichten' zum Vorschein gebracht würden, müßte bei solchen Ergebnissen geschlossen werdem, daß sich in den Antworten ein besonders 'mutiger' Mensch widerspiegele; jedenfalls könnte man mit Murray (1951) den in dem Testprotokoll (scheinbar) indizierten 'Mut' als 'a relevant constituent' des Probanden auffassen. Von der Voraussetzung ausgehend, daß stimuli mit hoher cue-Stärke diagnostisch deshalb besonders aufschlußreich seien, weil ihre Vermeidung auf ego-defense schließen läßt, müßte hier gefolgert werden, daß eine Verteidigungsreaktion nicht vorliege und daß bei diesen Vpn das Fallschirmspringen nicht konfliktbehaftet sei (was jedoch beides der Fall ist). Mit der Auffassung, daß

- im Gegensatz zu diesen (wenn auch stärker strukturierten) projektiven stimuli - ein self-report (nämlich 'eindeutige' stimuli) nur Oberflächen-Erscheinungen hervorzubringen vermöchte, kann nicht vereinbart werden, daß im Gegenteil in diesem Experiment bei direkter Befragung - der Realität entsprechend - die Angst nicht verschleiert werden kann. Freilich läßt sich eine Befragung im Zustande a k t u e l l e r Motiviertheit (Angst) und zugleich in einer Situation, die eben diesen Zustand determiniert (kurz vor dem Absprung), nicht vergleichen mit self-reports, bei denen in neutraler Situation allgemeine oder d i s p o s i t i o n e l l e motivationale Tendenzen erforscht werden sollen. Gleichwohl wird deutlich, daß projektive stimuli zu Auslösern auch von 'Kontrollfunktionen' werden können. Umgekehrt kann es bei eindeutigen stimuli eintreten, daß sich der eigene drive-state nicht mehr defensiv verbergen läßt. Es erweist sich auch hier als völlig unzulänglich, das projektiv Geäußerte 'als solches' bzw. von vorneherein als Ausdruck einer bestimmten 'Ebene' zu interpretieren. Handelt es sich um einen 'mutigen Sportler', um einen 'ängstlichen', aber diese Angst kompensierenden Fallschirmspringer? Ausschlaggebend ist schließlich - solange das Verhältnis von Index und Indiziertem nicht eindeutig ist - das V e r h a l t e n selbst als Resultante aus dem drive-state und den Verhaltensstandards : das overt behavior muß mit berücksichtigt werden, soll die Antriebsseite zureichend beurteilt werden können.

Nicht nur die Testreaktion läßt sich also 'als solche' nicht beurteilen, auch das Motiv selbst ist nicht 'als solches' schon hinreichend bestimmt. Es kann vielmehr auch von der mit ihm verknüpften 'Kontrolle' nicht losgelöst werden; denn es tritt nicht unabhängig von dieser in Erscheinung, sei es im realen - oder im Testverhalten. Diese Auffassung wird insbesondere bei Klein (1954) sehr präzise formuliert : "... we must wonder how 'pure need' can ever be tapped or how any single piece of behavior can be measured for so-called 'intensity' of need alone. Not only is the intensity of a need hard to gauge if we fail to take account of controlling mechanisms; we can hardly reach correct conclusions about

the quality of a need, its effects as an incitor of behavior, or even its presence" (S.228). "Need, than, is regulated - not necessarily out of existence - but into definite channels of expression and influence in keeping with the reality context." (S.234)

Man wird also nicht die 'Persönlichkeit' dadurch beschreiben müssen, daß sie aus mehreren 'Ebenen' von 'impulse-control systems' bestünde, wie Stone und Dellis annehmen (vgl. S.38f). Beide Tendenzen, impulse und control, sind vielmehr in den verschiedenen Formen ihrer Balance zu betrachten : die eine oder andere Tendenz wird in größerem oder geringerem Maße überwiegen, die Person wird im Sinne der einen o d e r der anderen r e a g i e r e n .

Die Arbeiten, welche vom Konflikt-Modell ausgehen, haben gezeigt, daß eine der Bedingungen, welche über das jeweilige Verhältnis beider Tendenzen entscheidet, im stimulus-Material selbst liegt. Tests verschiedener stimulus-Relevanz richten sich jedoch nicht (wie es im vagen Ebenen-Konzept angenommen wird, s.S.37ff) auf verschiedene 'Segmente' der Person, welche je einzeln in der Person 'vorliegen' und so je von verschiedenen Tests 'angesprochen' und abgebildet würden [1]. Stimuli verschiedener starker Relevanz rufen vielmehr verschiedenartige Reaktionen (nicht Segmente) hervor, bedingen ein je verschiedenes Verhältnis von impulse und control oder von need und regulation und fungieren so als - eine - Bedingung dafür, daß entweder ein Motiv defensiv umgangen, daß der Anspruch, es zu kontrollieren, formuliert oder daß es selbst - je nach der Art des Testmediums - in mehr oder weniger indirekter Form ausgedrückt wird.

Welche Reaktion jedoch tatsächlich erfolgt, hängt wesentlich von den Bedingungen der Motivkontrolle in der Person selbst und damit auch von der Art der sozialen Relevanz des Motivs ab : ist es ein gemeinhin sanktioniertes, ein sozial akzep-

---

1) Jene Autoren, welche annehmen, daß Tests sich danach unterscheiden, wie 'tief' die Segmente liegen, die sie abbilden, glauben, wie bereits dargestellt, die Funktionsweise der Verfahren durch den Ort auf einem 'Tiefenkontinuum' bestimmen zu können (vgl.S. 38).

tiertes Motiv, ist es ein realisiertes, ein unbefriedigtes, ein mit Schuld oder Angst besetztes Bedürfnis oder eine konfliktfreie Verhaltensdeterminante? Es genügt also nicht, nur formal von verschiedenen 'impulse-control-systems' zu sprechen und verschiedene Tests verschiedenen derartigen 'Systemen' zuzuordnen; sondern die 'Ebenen' sind inhaltlich so zu definieren, daß auch Begriffe wie 'befriedigtes - unbefriedigtes Bedürfnis', 'Konflikt', 'Angstbesetzung' in das Ebenen-Konzept aufgenommen werden können.

Diese persönlichkeitsdynamischen Beziehungen sind bisher für die Belange der Diagnostik am differenziertesten formuliert in der Theorie von Leary : hier jedoch wird das Problem der Funktionsweise projektiver Tests völlig unzureichend behandelt. Bei Rosenzweig fehlt es demgegenüber an einer Persönlichkeitstheorie. Beide Autoren vernachlässigen eine Analyse der Bedingungen der Testreaktion : es werden weder die Bedingungen auf Seiten des Testverfahrens, noch die differentiellen Bedingungen auf Seiten der Person (des Probanden) veranschlagt, geschweige denn untersucht. Eine solche Analyse wird im folgenden noch weiter vorzunehmen sein, damit die Rolle von approach- und avoidance-Tendenzen präziser gekennzeichnet und der Ansatzpunkt für ein Ebenen-Konzept näher bestimmt werden kann.

b) Zur Rolle von perceptual sensitization und defense

Die Anwendung des Konfliktmodells hat zu den bis heute differenziertesten Einsichten über das Verhältnis von Test-Index und Indiziertem geführt, indem es die Bedingungen zusammenhängend beschreibt, die der Relation zwischen Teststimulus auf der einen und Test-response auf der anderen Seite zugrundeliegen. Indem der Vorgang erkennbar wird, nach welchem ein bestimmter stimulus zum Auslöser für eine bestimmte Reaktion wird (Reizgeneralisation zwischen Test- und extra-Test-stimuli, Widerstreit zwischen approach- und avoidance-, bzw. zwischen expression- und inhibition-Tendenzen), ist ein wichtiger Ansatzpunkt für eine Interpretation des Test-Index' selbst gegeben.

Gleichwohl wird ein wesentlicher Aspekt bei diesen Arbeiten vernachlässigt. Es werden zwar die stimuli in ihren objektiven Bedingungen und auf Seiten des Subjekts die Motivstärke in Betracht gezogen; doch wird als vermittelnder Instanz zwischen beiden der W a h r n e h m u n g der stimuli keine eigene Bedeutung beigemessen. In diese Wahrnehmung aber gehen bereits persönlichkeitsspezifische Determinanten ein.

Insbesondere bei den Untersuchungen, in welchen nicht nur allgemein 2 widerstreitende Tendenzen angenommen werden (wie etwa bei Lesser, vgl.S.66ff), sondern in welchen das Konflikt-Modell in seinen experimentellen Anwendungsmöglichkeiten durch systematische Bedingungsvariation ausgeschöpft werden soll, kommt der Rolle der stimulus-Gegebenheiten ein einseitiger Vorrang zu. Sind cue-Stärke und Motivstärke gegeben, so ergibt sich - diesem Modell nach - als deren Funktion notwendig ein bestimmter Motivausdruck.

Der Begriff der Reizgeneralisation kann jedoch hier, im Bereich der Reaktion auf projektive stimuli, nicht als ein Begriff der objektiv feststellbaren Ähnlichkeit zwischen 2 Reizgegebenheiten gefaßt werden, also als ein Begriff der 'experimentellen Ökologie' [1], sondern nur als ein 'Lebensraum-Begriff' [2], welcher einen subjektiven Vorgang, nämlich die subjektive Wahrnehmungsleistung, also eine subjektiv e r l e b t e Ähnlichkeit beschreibt. Entsprechend argumentiert z.B. Klein (1954), daß response-Generalisierung nicht durch gleichartige physikalische stimuli, sondern durch solche Situationen, welche für das Subjekt ähnliche adaptative Probleme stellen, hervorgerufen werde. Die 'Anfangsbedingungen' sind nicht die stimuli in ihrer physikalischen Beschaf-

1) Der Begriff der 'experimentellen Ökologie' erfaßt die experimentelle Situation unter dem Aspekt der 'Welt für uns alle' und ist darin das objektive Korrespondenz des Begriffs des Lebensraumes, welcher die 'Welt für jeden einzelnen' bezeichnet. Ansatzpunkt für diese an Lewin (1935) orientierte Unterscheidung ist die für die Planung und Bewertung von psychol. Experimenten sehr wichtige Tatsache, daß "die Vp nicht etwa zwangsläufig das experimentelle Arrangement so erleben muß, wie es vom Experimentator als 'Ökologie' hergestellt ist" (Holzkamp 1964, S.171)

2) vgl. Fußnote 1)

fenheit, sondern "the expectations about the functions that a response must serve in a situation" (S. 267). Diese besondere Relation aber zwischen objektivem stimulus und persönlichkeits-determinierter Wahrnehmung wird erst mit den Begriffen der social-perception-Forschung hinreichend erfaßt.

Die Annahme einer festen Verknüpfung von Reiz und Reaktion (S - R) wird von diesem Ansatz zurückgewiesen. Wahrnehmung wird vielmehr durch das Prinzip der Selektivität bestimmt, welches zu inter-individuellen Differenzen in der Auffassung der stimuli führt. Aus dem Reizangebot werden jene stimuli selegiert, welche am stärksten den subjektiven 'Hypothesen' oder 'Schemata' entsprechen : dabei werden 'positive' stimuli noch subjektiv verstärkt, während unangenehme Reize abgeschwächt werden ('minimalization - maximalization'). Solcher Abwehr unangenehmer Reize (defense) steht jedoch das realitäts-orientierte Prinzip erhöhter Wachsamkeit bedrohlichen stimuli gegenüber (vigilance). Entweder sind die subjektiven Wahrnehmungshypothesen für bedrohliche Inhalte gering, so daß andere, stärkere 'Hypothesen' durch das Material bekräftigt werden [1]; oder aber es bestehen starke 'negative Hypothesen' (Postman 1951), so daß auch bei mehrdeutigen Reizgegebenheiten gerade Bedrohliches besonders sensitiv wahrgenommen wird.

Für die Belange der Diagnostik interessiert nun insbesondere die Frage, ob Personen etwa durchgehend stärker durch avoidance-Verhalten oder mehr durch sensitization charakterisiert sind. Erikson (1952) z.B. nimmt eine solche durchgängige Reaktionstendenz an, die sich gleichartig auf die Wahrnehmung wie

---

1) Die Frage, ob perceptual defense ein 'vorbewußter Akt' ist, also der kognitiven Wahrnehmungsleistung selbst vorausgeht, ob es sich dabei um das unterschiedliche Gewicht der positiven und negativen Hypothesen handelt oder aber nur darum, daß bei den entsprechenden Experimenten allein die Verbalisierung des Wahrgenommenen defensiv hinausgezögert wird (vgl. Whittacker 1952, Erikson, Browne 1956), ist in diesem Zusammenhang nur von sekundärem Interesse und kann hier nicht entschieden werden.

auf das Behalten auswirkt [1]. Diese Frage führt zu der hypothetischen Gegenüberstellung von 'sensitizers' und 'avoiders'.

Das Konfliktmodell erschließt demgegenüber nur die eine Möglichkeit der defensiven Reaktion auf ich-bedrohliche stimuli : je größer die stimulus-Relevanz für ich-bedrohliche Variablen, um so stärker die Verteidigung gegen sie. Warum stattdessen bei geringer stimulus-Relevanz ich-bedrohliche cues (welche dann 'vermieden' werden) überhaupt wahrgenommen werden, vermag es nicht zu klären. Ebensowenig ist diesem Modell zu entnehmen, daß ein bedrohlicher content-pull nicht umgangen, sondern 'realitätsgerecht' auf ihn reagiert wird, wie es in den Experimenten von Epstein (vgl.S.68ff) schon deutlich wurde. Bei Epstein spiegelte diese Reaktion den Versuch wider, gegen die Ich-Bedrohung (Angst) anzugehen, statt sich ihrer zu entziehen. Bisher fehlt allerdings eine systematische Analyse zu dieser Frage zweier entgegengesetzter Reaktionsarten. Die wenigen Untersuchungen, die eine sensitization nachzuweisen versuchen, bieten - ihrer methodischen Schwächen wegen - keine hinreichende Beurteilungsgrundlage.

So konnte etwa die Erwartung, daß Selbsteinsicht in die eigene Bedürfnislage (hostility) zu sensitization und fehlende Selbsteinsicht zu defense führen würde, nur teilweise (und nicht überzeugend [2]) bestätigt werden : der Rorschach erwies sich als unempfindlich für diesen Unterschied, nur bei einem Preference-Test konnten teilweise, der Erwartung entsprechend, sign. Unterschiede nachgewiesen werden (Abrams 1962). Aufschlußreicher ist eine Arbeit von Phares (1961), nach welcher besonders ängstliche Vpn (gemessen durch die Taylor-anxiety-scale) mit sign. mehr bedrohlichen Geschichten auf

1) Diese Annahme wird in der Literatur unter der Bezeichnung 'constance hypothesis' diskutiert und ist selbstverständlich nicht mit der 'Konstanzannahme' der älteren Psychophysik zu verwechseln.

2) Problematisch ist hier die Messung der Variable 'Selbsteinsicht' : wird fehlende Selbsteinsicht nämlich durch die Differenz von Selbst- und Fremdbeurteilung definiert, so ist sie nicht unabhängig von der Variablen'Ich-Verteidigung', mit welcher sie ja in Beziehung gebracht, als deren 'Bedingung' sie verstanden werden soll. Schätzt eine Vp sich als weniger hostile ein als andere sie, so kann schon dies - wider bessere, also tatsächlich vorhandene Selbsteinsicht - durch eine Verteidigungshaltung bedingt

den TAT reagieren als nicht-ängstliche Vpn. Zwar handelt es sich hier nicht um die freie Produktion von Geschichten, sondern um die rangmäßige Zuordnung von v o r g e l e g t e n Geschichten (mit neutralem und bedrohlichem Gehalt) zu den TAT-Karten. Doch ist jedenfalls - für diese besondere Art der Test-Beantwortung ('Reaktionstyp' [1]) - nachgewiesen, daß ängstliche Vpn besonders sensitiv auf den bedrohlichen Gehalt der Geschichten reagieren und ihn nicht von sich abwehren, wie es dem Konfliktmodell nach zu erwarten wäre. Allerdings ist ein solches Ergebnis beschränkt auf die besondere Art des geforderten Reaktionstyps (vgl.Fußnote 1) und nicht übertragbar auf die freie Produktion von Geschichten, in welchen wahrscheinlich auch diese Vpn die Verbalisierung (expression) bedrohlicher Inhalte vermieden hätten. [2] Doch verweist es darauf, daß - ungeachtet des stimulus-Materials - auch der geforderte Reaktionstyp ausschlaggebend zu sein scheint, ob mit sensitization oder mit defense reagiert wird, ausschlaggebend also für das jeweilige Verhältnis von Test-Index und Indiziertem. Diese Annahme, daß sensitization und defense nicht als konsistente Persönlichkeitstendenzen aufgefaßt werden können, sondern von den Bedingungen des Tests

---

sein. Somit wäre anstelle der hier mit falschen Mitteln geprüften Hypothese zu untersuchen, ob 'tatsächliche' Selbsteinsicht mit defense wie mit sensitization gepaart sein kann.

1) Dieser Begriff bezeichnet die bei einem Test geforderte Reaktionsart : nicht subjektive Eigenarten der Reaktion sind also gemeint (wie z.B. defense oder sensitization), sondern die Unterschiede zwischen 'Geschichte-erzählen', 'Sätze-ergänzen', 'Wünsche-äußern', etc. vgl.Fußnote 1).S.13.

2) Diese Arbeit kann aus diesen methodischen Gründen nicht, wie Phares es unternimmt, den Arbeiten von Lesser gegenübergestellt werden, welcher bei Aggressions-Angst gerade die Vermeidung aggressiver cues bei TAT-ähnlichen Bildern nachwies, nämlich hier - bei 'freier' Produktion - eine 'thematische' Vermeidung aufzeigte.

wie der untersuchten Variablen abhängen, wird hier auch dadurch bekräftigt, daß die Gruppe mit sign. mehr bedrohlichen Geschichten auf einen Satzergänzungstest nicht einheitlich, sondern teils mit avoidance (gegenüber den sexuellen und aggressiven Implikationen der Satzstämme), teils mit sensitization reagiert [1].

Diese Arbeiten lassen noch keine Entscheidung darüber zu, aufgrund welcher zusätzlicher Bedingungen mit sensitization oder defense zu rechnen ist. Sie machen jedoch deutlich, daß - entgegen der Konsequenz des Konfliktmodells - auch bei ich-bedrohlichen Inhalten mit derem Ausdruck, statt allein mit deren Vermeidung gerechnet werden kann. Der Motivausdruck verweist in solchem Falle nicht darauf, daß nur eine geringe oder keine avoidance-Tendenz und somit auch kein Konflikt vorliege; sondern hier wird indiziert, daß gerade die durch die cues ausgelöste Ich-Bedrohung zu deren sensitivierter Wahrnehmung führte.

In diesem Zusammenhang ist noch einmal auf die Arbeiten von Lesser (vgl.S. 66ff) und Mussen, Naylor (1954) zu verweisen. Auch hier führte die Aggressions-Angst nicht zur vollständigen Vermeidung aggressiver Inhalte in den Testreaktionen, sondern - so läßt es sich nun formulieren - es wurden aggressive cues besonders sensitiv wahrgenommen (hohe positive Aggressions-Indizes), aber sodann im nachhinein 'thematisch' vermieden. Daß aber beide Tendenzen, approach und avoidance, gleichzeitig die Testreaktion charakterisieren, läßt sich nur mit Hilfe der Begriffe 'sensitization' und 'defense' theoretisch begründen. Im Konflikt-Modell dagegen wird die Reaktion nur als R e s u l t a n t e aus beiden Tendenzen festgestellt : entweder die approach- oder die avoidance-Tendenz überwiegt und bestimmt das Verhalten.

Es stehen sich nunmehr folgende Formen der Reaktion gegenüber : die ich-bedrohlichen cues werden nicht wahrgenommen, weil andere 'Hypothesen' stärker sind; die cues werden besonders 'sensitiv' wahrgenommen : es wird mit 'avoiding'

1) vgl. dagegen die Konstanzhypothese von Erikson (1952), s.S.76f.

oder mit 'coping' reagiert; das Motiv selbst gelangt (substitutiv) zum Ausdruck, um dann durch 'thematische Vermeidung' wieder eingeschränkt zu werden; dieselben cues sind für bestimmte Vpn nicht bedrohlich : es wird der eigenen Motivstärke, bzw. dem offenen Verhalten entsprechend reagiert.

Es ist nun hinlänglich deutlich geworden, welche Mehrdeutigkeit einem Test-score anhaftet. Die einfache S-R-theoretische Zuordnung von stimulus-relevance, Motiv-Stärke und Motiv-Ausdruck erweist sich angesichts der tatsächlichen Vielfalt der Reaktionen und ihrer Bedingungen als unzulänglich. Populationen müssen nicht nur danach unterschieden werden, ob bei ihnen eine 'guilt-reducing-inhibition' (z.B. Aggressions-Angst, vgl. S. 67) vorliegt oder nicht, sondern es müssen offensichtlich sehr viel mehr Differenzierungen eingeführt werden. Welche Ansatzpunkte aber gibt es, derartige Differenzkriterien aufzudecken? Hier liegt eine Fülle von Problemen, für deren Lösung man den bisher vorliegenden Untersuchungen noch nicht genügend Hinweise entnehmen kann : auch solche Tendenzen wie z.B. sensitization oder defense scheinen eine Vp nicht durchgehend zu charakterisieren und können daher nicht als ein allgemeines Kriterium dafür dienen, mit welcher Art des Testverhaltens bei einer bestimmten Vp (resp. bei einer Gruppe von Vpn) zu rechnen ist. Entsprechendes stellten andere Autoren fest, nach welchen ähnliche Tendenzen, die hier als revealing- bzw. conceiling response style bezeichnet werden, ebenfalls nicht konsistent alle Reaktionen einer Vp bestimmen (Zimbardo 1964; Sherwood 1966).

Für die Ebenen-Diskussion ergibt sich angesichts dieses Materials, daß ein solches Konzept sich vor allem daran zu bewähren hat, ob es zu derartigen Differenzkriterien führt. Es wird allerdings - wenn überhaupt - nur einen Ausschnitt aus der denkbaren Fülle solcher Kriterien erbringen können (dazu ausführlicher später).

Abschließend sei noch einmal auf den Vorschlag von Farberow (1959) verwiesen, für jede einzelne Variable ein nach dem Grade der stimulus-Relevanz abgestuftes stimulus-Material

zu bieten. Erst bei einer solchen systematischen Gegenüberstellung kann - u.U. - präziser festgestellt werden, welche Inhalte bereits bei stimulus-Mehrdeutigkeit (also ohne entsprechenden content-pull und nur aufgrund der Stärke der subjektiven Hypothesen) hervorgebracht werden; welche Inhalte auf der Grundlage von Motiv-relevanten stimuli v e r m i e d e n werden; welche cue-Stärken also nötig sind, um ein Motiv zum Ausdruck zu bringen oder dessen Vermeidung zu bewirken. [1] Es wäre damit die bisher differenzierteste Methode entwickelt worden, die relative Stärke eines Motivs und der Kontrolle seines Ausdrucks zu bestimmen. Allerdings wird auch hier - als Grundlage der Validierung - vorausgesetzt, daß bei bestimmten stimulus-Gegebenheiten bestimmte Motivbedingungen notwendig zu einem bestimmten Motiv-Index führen. Zusätzliche Bedingungen, wie z.B. sensitization, coping, revealing response style oder bestimmte Rollenerwartungen (vgl.S.57f) werden auch hier nicht berücksichtigt. Die Validierung aber auch eines solchen Verfahrens mit systematisch variierten stimulus-Bedingungen kann nur für verschiedene Gruppen getrennt vorgenommen werden.

4. Folgerungen aus den zitierten Untersuchungen

Der Überblick über die verschiedenen Forschungsansätze und Ergebnisse zur Frage des Motivausdrucks bei projektiven Verfahren und damit die Übersicht über dessen vielfältige Bedingungen läßt die Farberow'sche Konsequenz zumindest als eine fruchtbare Alternative zu einem vagen Ebenen-Konzept erscheinen. Es werden hier nicht, wie bei der vagen Verwendung des Ebenen-Begriffs, die Tests durch die Annahme relativiert, daß sie - entlang eines (imaginären) Tiefenkontinuums - in nicht näher feststellbarer Weise Verschiedenes messen, - was dann begrifflich als verschiedene layers, stratas oder segments bzw. levels bestimmt wird. Hier werden vielmehr Tests gerade daraufhin konzipiert, daß sie in systematischer Abstufung der stimulus-Relevanz ein und dasselbe

1) vgl. hierzu auch das sehr detaillierte Programm bei Heath (1958).

Phänomen (das betreffende Motiv) sozusagen auf verschiedenen Wegen anzielen und damit erst das jeweilige Verhältnis von impulse und control näher bestimmen lassen. Nicht um ein 'impulse-control-system' und die ihm zugehörige 'Ebene' (s.S.38f) geht es hier, sondern um das jeweilige Verhältnis zweier Tendenzen (vgl. need und regulation). Die Annahme verschiedener zwischen 'Oberfläche' und 'Tiefe' aufeinandergeschichteter Segmente erübrigt sich : denn der Motivausdruck kann in seinen Bedingungen, nämlich durch einen Hinweisreiz ausgelöst zu werden und in verschiedenen Formen von defense oder Bewußtseinskontrolle zurückgehalten, modifiziert oder auch 'freigelassen' zu werden, sehr viel differenzierter beschrieben werden.

Erst bei einer solchen Gegenüberstellung der Ergebnisse der allgemein-psychologisch orientierten Untersuchungen projektiver Tests mit den am Ebenen-Konzept und damit mehr persönlichkeitstheoretisch orientierten Arbeiten drängt sich die kritische Frage auf, ob nicht auf das (bisher unergiebige) Ebenen-Konzept verzichtet werden müsse. Vergleicht man den Ertrag, den beide Richtungen bis heute erbrachten, so kann diese Frage uneingeschränkt bejaht werden. Anderes gilt jedoch für eine Anwendung der Forschungsergebnisse auf die Belange der praktischen Diagnostik. Hier besteht nicht allein das Problem, daß es einen solchen differenzierten, nämlich auf der Grundlage verschiedener stimulus-Stärken validierten Zugang zu einzelnen Persönlichkeitsvariablen bisher nicht gibt. Hier ist die weitere Aufgabe gestellt, die allgemeinpsychologischen Begriffe wie approach-avoidance, sensitization-defense, expression-inhibition zurückzuführen in einen persönlichkeitstheoretischen Zusammenhang und damit in eine Sprache zu übersetzen, welche dem diagnostischen Ziel einer Persönlichkeitsbeschreibung entspricht. [1)]

---

1) Wenn hier die allgemeinpsychologischen Begriffe und die vermittels solcher Begriffe gewonnenen Erkenntnisse in das Sprachsystem eines persönlichkeitstheoretischen Modells übersetzt werden sollen, so impliziert ein solches Vorgehen Grundannahmen sowohl über die Aufgaben einer diagnostischen Persönlichkeitsbeschreibung, wie auch über die gegenseitigen Beziehungen derartiger Sprachsysteme.

Die Aufgabe der Diagnostik soll hier im weitesten Sinne verstanden werden. Hörmann charakterisiert sie allgemein

Welche Rolle spielt beispielsweise das theoretische Konzept 'conflict' (hier verstanden als Konflikt zwischen approach- und avoidance-Tendenz), wenn die Aufgabe einer diagnostischen Persönlichkeitsbeschreibung gestellt ist? Diente dieser Begriff 'conflict' bei experimentellen Analysen zur Bezeichnung einer Bedingung für die Art des Motiv- A u s d r u c k s in der Testreaktion, so hat der Diagnostiker zu fragen, welches die Bedingungen für diesen 'conflict' selbst sind. Er, der Diagnostiker, fragt von einem Gesichtswinkel aus, für welchen die Testreaktion nicht Gegenstand (und Zweck) der Analyse ist, sondern Mittel dafür, Kenntnisse über eine Person zu gewinnen. Ihn interessieren das alltägliche Verhalten und d e s s e n Bedingungen. Ein Bedürfnis muß für ihn beispielsweise danach beschrieben werden können, ob es an seiner verhaltensmäßigen Realisation gehindert werden muß, weil es nicht der sozialen Norm entspricht (z.B. aggression, hostility), ob es selbst oder seine verbale

---

genug als "Hilfe bei der Orientierung in der sozialen Umwelt" (1964b, S.7). Versteht man diesen Begriff der 'Orientierung' im weitesten Sinne, so hat es die Diagnostik - zumindest a u c h - mit spezifisch i n d i v i d u e l l e n Problemen zu tun. Damit aber steht die Diagnostik vor Aufgaben, die über jene partiellen Fragen hinausgehen, welche bestimmte Institutionen der Gesellschaft (seien es Schule, Industrie, Verwaltung oder Militär) an sie richtet. Sie steht vor Aufgaben, welche mit den Mitteln der allgemeinen, wie der differentiellen Psychologie (letztere ist eine Abart der ersteren, vgl. etwa Allport 1949) allein nicht zu lösen sind, Aufgaben also, bei denen es nicht allein darauf ankommt, dem Individuum einen Ort innerhalb der Häufigkeitsverteilung eines isolierten Merkmals zuzuweisen.

Persönlichkeitsbeschreibung im weitesten Sinne setzt eine Theorie voraus, welche die 'Persönlichkeit' in Kategorien s o z i a l e n Verhaltens und seiner Bedingungen erfaßt und welche zugleich Begriffe liefert, die sich für die K o m m u n i k a t i o n zwischen Diagnostiker und Proband eignen. Denn 'Hilfe' ist ohne Kommunikation nicht denkbar und 'Orientierungshilfe' nicht ohne Analyse des Verhaltens in seinen sozialen Bestimmungen. Eine solche Theorie wird ohne Bezugnahme auf i n t e r individuelle Differenzen nicht ein vollständiges Instrumentarium für die Persönlichkeitsbeurteilung bereitstellen können; ohne einen Begriffsapparat zur Beschreibung der i n t r a - individuellen Organisation aber wird sie ebenfalls ein unzureichendes Mittel sein. Eine Theorie, welche diese intra-individuelle Dimension als

Äußerung aufgrund kultureller Norm von Schuldgefühlen begleitet ist (z.B. sex), ob im Falle eines sozial positiv bewerteten Verhaltens die antriebsmäßige Seite stark genug ist, den Verhaltensanspruch zu verwirklichen oder ob hier der Verhaltensanspruch selbst gering ist (z.B. achievement) oder ob schließlich das Verhalten dem eigenen Anspruch und zugleich auch den motivationalen Tendenzen entspricht : sei es hier, daß ein sozial-erstrebenswertes Verhalten stark ausgebildet wurde (z.B. nurturance) und ein abzulehnendes Motiv gar nicht nach seiner Verwirklichung drängt (z.B. aggression) oder daß umgekehrt ein subjektiver Anspruch nicht besteht, bestimmte positive Normen zu erfüllen oder ein gemeinhin negativ bewertetes Verhalten unter Kontrolle zu bringen.

Bezugsrahmen mit umfaßt, ist - eine Persönlichkeitstheorie.

Wir können das Verhältnis von idiographischer Beschreibung und nomothetischer Wissenschaft hier nicht im einzelnen diskutieren. Doch sei auf das Verhältnis zwischen persönlichkeitstheoretischem und allgemeinpsychologischem Kategoriensystem im Rahmen unserer Fragestellung kurz eingegangen.

Die allgemeine Psychologie hat im Bereich der Diagnostik die entscheidende Funktion, jene Variablen zu isolieren, welche als Maßdimensionen für inter-individuelle Differenzen herangezogen werden können. Diese Variablen jedoch und damit die Information, die sich (experimentell gesichert) einem Test entnehmen läßt, können - für den Zweck einer Persönlichkeitsdiagnostik im weitesten Sinne - nur dann in ihrem Aussagewert beurteilt und praktisch verwendet werden, wenn sie sich in Beziehung bringen lassen zu jenen Kategorien, welche 'soziales Verhalten' und 'intra-individuelle Zusammenhänge' erfassen.

Eine derartige Übertragung allerdings eines Sprachsystems in ein anderes impliziert, daß diesen Systemen nicht die Funktion des 'Abbildens', sondern nur die des 'Ordnens', eines Ordnens also von verschiedenen Gesichtspunkten aus, zukommt. Sie impliziert in diesem Falle ferner, daß das persönlichkeitstheoretische System eine weniger direkte bzw. weniger geprüfte 'Realitätsbeziehung' und damit einen größeren 'Bedeutungsüberhang' (vgl. Holzkamp 1964) aufweist.

Eine solche Persönlichkeitstheorie, welcher experimentell gesicherte (allgemeinpsychologische) Erkenntnisse zugrundeliegen, welche orientiert wäre an der Diagnostik im weitesten Sinne und zugleich Maßdimensionen für die durch sie definierten Variablen enthielte, gibt es bis heute nicht. Richtungweisend in der Verknüpfung von Persönlichkeitstheorie und Diagnostik war derzeit das System von Murray (1939). Ansätze von einer allgemeinpsycholo-

Gerade diese für eine Diagnose erst aufschlußreichen Beziehungen zwischen Antrieb, seiner Kontrolle und dem jeweils resultierenden Verhalten können durch jene allgemeinpsychologischen Modelle nicht hinreichend erfaßt werden, welche nur S-R-Verknüpfungen zur Grundlage haben oder allein am expectancy-Begriff oder allein wahrnehmungstheoretisch orientiert sind. Hier ist vielmehr ein Persönlichkeitsmodell nötig, welches die Test-Information in einen dynamischen Zusammenhang stellt, von welchem her erst Aussagen über die Person, und zwar dies über die besondere Testsituation hinaus, gemacht werden können, - seien sie auch noch so spezifisch nur auf einzelne Persönlichkeitsvariablen beschränkt.

Hier sei eingeflochten, was an späterer Stelle (s.S.116ff) noch ausführlicher erörtert wird : daß nämlich experimentelles Vorgehen sich nur auf einzelne, exakt bestimmbare Variablen richten kann; denn anders erfüllte es nicht die ihm - ex definitione - zugesprochene Funktion der K o n t r o l l e . Nicht die 'Ganzheit' der Person, nicht ihre 'charakterologische Struktur' ist hier gemeint, wenn ein Persönlichkeitsmodell gefordert wird.

Wohl aber wird ein Modell benötigt, in welchem einzelne Persönlichkeitsvariablen sich in einen persönlichkeitstheoretischen Zusammenhang stellen lassen. Eben dieses leistet ein Ebenen-Konzept, wenn es - wie bei Leary - als eine Persönlichkeitstheorie formuliert wird. Hier wird eine 'Eigen-

---

gisch orientierten Position aus finden sich heute in der Schule von Klein.

In dieser Arbeit kann - für den Gegenstandsbereich projektiver Methoden - nur ein erster Schritt von allgemeinpsychologischen Begriffen zu persönlichkeitsorientierten Konzepten gegangen werden. Das Ebenen-Konzept ist weit davon entfernt, eine Persönlichkeitstheorie zu liefern. Es ist ausgerichtet nur darauf, Test-Reaktionen, soweit sie allgemeinpsychologisch analysiert wurden, auf der Ebene persönlichkeitstheoretischer Zusammenhänge zu beschreiben und damit Begriffe zu liefern, die sich zu einer der diagnostischen Aufgabe angemessenen, wie experimentell fundierten Interpretation von (projektiven) Test-scores eignen.

schaft' - bis heute zentraler Begriff in der Persönlichkeitsdiagnostik - nicht als ein reales 'Agens', nicht als eine Konstante des 'Charakters' mißverstanden [1]. Sondern hier wird der Begriff 'Eigenschaft' als eine theoretische Verhaltensvariable konzipiert, welche erst hinlänglich bestimmt ist, wenn sie zugleich auf allen Ebenen des Modells erfaßt wird. Motiv, Verhaltensanspruch, Motivkontrolle und resultierendes Verhalten sind an ein und derselben Verhaltensvariable zu untersuchen; oder : eine 'Eigenschaft' ist - wie wir es nennen wollen - in ihren verschiedenen 'Erscheinungsmodi' zu untersuchen.

Eine solche Persönlichkeitstheorie aber leistet ein weiteres. Sie liefert nicht nur das Begriffsgerüst, die 'Erscheinungsmodi einer Eigenschaft' und somit verschiedene Testergebnisse (ihre Gültigkeit vorausgesetzt) in ihrem gegenseitigen Zusammenhang zu erfassen. Sie beschreibt vielmehr auch jene Bedingungen, welche der Testreaktion v o r a u s g e h e n : jene Bedingungen somit, welche - auf Seiten des Subjekts - die Art der Testreaktion bestimmen. Aggressive Tendenzen z.B. kommen um so weniger zum Ausdruck, je mehr sie in der Erziehung sanktioniert wurden, je mehr sie also - dem subjektiven Anspruch nach - unter Kontrolle gehalten werden müssen. Das Verhältnis zwischen Ich-Ideal (Ebene V bei Leary, vgl.S. 48) und Verhalten, bzw. zwischen Ich-Ideal und Motiv wird also beispielsweise nicht nur darüber entscheiden, ob aggressive Tendenzen im o f f e n e n Verhalten beobachtbar werden, sondern auch darüber, ob und wie sie als 'latente Inhalte' durch T e s t methoden, resp. projektive Tests, sichtbar gemacht werden können.

Die Mehrdeutigkeit von Test-scores, die sich nach dem Prinzip einer empirischen Validierung (Korrelation von Test-score und (zumeist) offenem Verhalten) als eine ungenügende Gültigkeit des Tests darstellt, ist daher nicht dem Instrument anzu-

1) vgl. die Diskussion des Eigenschaftsbegriffs bei Hörmann (1964b) und seine Kritik an dem naiven Begriffsrealismus, 'Eigenschaften' als real Existentes, anstatt als theoretische (bewußtseinsimmanente) Konstruktion aufzufassen.

lasten, sondern den je verschiedenen Bedingungen, welche - auf Seiten der Person - zu der Reaktion führten. Liegen aber den Testreaktionen Bedingungen zugrunde, welche für das Verständnis dieser Reaktionen entscheidend sind, ohne direkt aus den Testwerten ablesbar zu sein (was Heiß 1951 die 'Mehrfunktionalität des Zeichencharakters' nennt), so sind damit besondere Schwierigkeiten für eine Validierung gegeben. [1)]

Eine direkte Punkt-zu-Punkt-Validierung etwa der Form 'hohe Aggressions-scores = hohe offene Aggressivität', also der direkte Schluß von Verhalten auf Verhalten, ist nicht möglich. Vielmehr muß ein Validierungsmodell herangezogen werden, anhand dessen sowohl die Bedingungen des zu diagnostizierenden V e r h a l t e n s , wie die (durch dieselben Bedingungen mitbestimmte) Genese des T e s t wertes mit in Betracht gezogen werden können. Ein solches Validierungsmodell liefert die construct-Validierung, wie sie von Cronbach und Meehl 1955 entwickelt wurde. Dieses Modell zeichnet sich dadurch aus, daß es auch die Analyse jener Bedingungen mit umfaßt, welche der empirisch feststellbaren Korrelation zwischen Testverhalten und Kriteriumsverhalten (also der empirischen Gültigkeit) zugrundeliegen. Die construct-Validierung setzt, im Gegensatz zur empirischen Validierung, eine Theorie voraus. Eine solche Theorie muß es leisten, z.B. eine 'Eigenschaft' so zu definieren, daß sich Hypothesen ableiten lassen über das durch diese Eigenschaft benannte Verhalten, sowohl in der T e s t -, wie in der K r i t e r i u m s situation.

Indem eine solche Theorie die Bedingungen beschreibt, die - bei gegebenen stimuli - auf Seiten des Probanden der einen oder anderen Reaktionsart zugrundeliegen, liefert sie auch die theoretische Grundlage für eine differentielle Validierung : sie liefert ein Begriffssystem, mit welchem sich die Differenz - K r i t e r i e n formulieren lassen, jene Kriterien, nach welchen Gruppen mit unterschiedlicher Testreaktion sich systematisch unterscheiden und so auch getrennt für Validierungsuntersuchungen heranziehen lassen.

---

1) Nach Heiß (1951) ist diese Mehrdeutigkeit, die er als 'Mehrfunktionalität des Zeichencharakters' bezeichnet (ein

Auf solchem Wege einer differentiellen Validierung könnte etwa festgestellt werden, daß der Test bei der Gruppe A in der Form 'a' funktioniert, während er bei Gruppe B nach dem Schema 'b' interpretiert werden müßte. Der Test wäre nun nicht mehr als 'mehrdeutig' abzuwerten, wie es sich ergeben muß, wenn man die verschiedensten Vpn einer Validierung zugrundelegt, in der falschen Annahme, die Vpn gehörten der gleichen Merkmalsklasse an.

Eine solche 'differentielle Validität', gewonnen nach dem Prinzip der construct-Validierung, entspräche schließlich einer 'empirischen Validität' auf 'höherem theoretischem Niveau' : denn sie implizierte jene Differenzkriterien, welche erst eine zuverlässige e m p i r i s c h e Zuordnung von Index und Indiziertem überhaupt ermöglichen.

Zu fragen ist nun, ob insbesondere das Ebenen-Konzept einen derartigen theoretischen Rahmen für eine construct-Validierung und damit Kriterien für eine differentielle Validität liefern kann. Nur wenn entschieden ist, ob es diese Forderung erfüllt, kann auch entschieden werden, ob die Ebenen-Diskussion im gegenwärtigen Stadium weiterführend ist, weiterführend insbesondere darin, daß die Ergebnisse der verschiedenen stärker allgemeinpsychologisch orientierten Experimente eine gemeinsame Interpretationsgrundlage erhalten, sich also in einen weiteren theoretischen Zusammenhang integrieren lassen.

---

manifester Zeichengehalt kann mehrere latente Zeichengehalte enthalten), allerdings ein positives Spezifikum der projektiven Tests (hier 'Entfaltungstests'), denn auf dem (nicht irgend präzisierten) Wege der 'D e u t u n g' läßt sich hier, im Gegensatz zu psychometrischen Verfahren (hier 'Prüftests'), das 'Zeichengesamt' eines individuellen Testprotokolls mit dem 'Strukturgesamt' der Persönlichkeit gleichsetzen : "Solange wir glauben, daß in einer Einzeläußerung oder einer Einzelhandlung die Persönlichkeit oder auch nur ein Teil von ihr erscheint, ist auch die Annahme, daß in einem Äußerungsgesamt die Persönlichkeit sich auswirkt und erscheint, durchaus berechtigt" (S.17). Der Unterschied zwischen dem Prinzip der differentiellen Validität und dem der erfahrungsgeleiteten 'Deutung' des Klinikers braucht kaum eigens hervorgehoben zu werden.

Das Ebenen-Konzept bietet sich an, weil hier die gleichen Bedingungen für das Verhalten verantwortlich gemacht werden, welche - allem Anschein (nämlich den hier aufgeführten Experimenten) nach - auch als Bestimmungen für die Art des Motivausdrucks in der Reaktion auf projektive Tests fungieren können.

Für die Planung eines Experiments ergibt sich aus dieser Diskussion des Begriffs der differentiellen Validität, daß sich für ein projektives Verfahren allgemein nachweisen lassen muß, daß es bei Vpn, die sich nach bestimmten Kriterien unterscheiden, seinerseits unterschiedlich 'funktioniert'. In Begriffen des Ebenen-Konzepts gesprochen, muß sich zeigen lassen, daß ein Test bei Vpn, die sich in ihrem 'Ebenen-Gefüge' unterscheiden, unterschiedliche R e a k t i o n e n hervorruft. Das Ebenen-Konzept ist daran zu b e w ä h r e n, ob es zu jenen Kriterien führt, welche Gruppen systematisch unterschiedlicher Reaktionen konstituieren. Das Ziel der experimentellen Analyse ist es also, T r e n n k r i t e r i e n für eine (zukünftige) differentielle Validierung aufzufinden.

Dieser theoretische Ansatzpunkt ist unvereinbar mit jenem vagen Ebenen-Begriff, wie er im vorangehenden Kapitel erörtert wurde (vgl. S.37ff). Die fehlende empirische Validität eines projektiven Verfahrens kann man nicht länger durch den Hinweis 'wegdiskutieren', daß sich in dem Test ein anderes 'Segment' der Person manifestiere als in dem gemessenen Kriteriumsverhalten. Verschiedene Tests sprechen nicht - ihrer unvereinbaren Ergebnisse wegen - 'Segmente verschiedener Tiefe' an. Bestimmte Bedingungen auf Seiten der Person (der besondere 'Kontext' des Motivs, ausgedrückt in Ebenen-Begriffen) geben - auch bei e i n e m Test - den Ausschlag, wie stark die approach- oder avoidance-Tendenz für den Ausdruck dieses Motivs ist. Die approach-Reaktion kann sich dabei als Ausdruck der motivationalen Tendenz selbst oder als Ausdruck des Versuchs, diese zu steuern (coping), vollziehen. Der tatsächliche Motivausdruck (expression) kann in diesem Sinne dann verstanden werden als verschiedene

'Ebenen' repräsentierend. Damit wird es möglich, von den verschiedenen 'Funktionsebenen' eines Tests zu sprechen, - wie Rosenzweig es als Programm aufstellte. [1)]

Ein solches Konzept muß jedoch zunächst erst entwickelt werden : es werden einzelne 'Persönlichkeitsebenen' so konzipiert werden müssen, daß deren 'Zueinander', deren 'Gefüge' sich in Beziehung bringen läßt zu den verschiedenartigen Reaktionsformen auf projektive Tests, wie sie in obigen Experimenten nachgewiesen wurden. Ein solches Konzept wird daher eine Art 'Kompromiß' darstellen zwischen dem Ansatz von Leary (also dem Begriff der 'Persönlichkeitsebene') und dem von Rosenzweig (also dem Begriff der 'Reaktionsebene') : das Modell der Persönlichkeit soll zugleich Ansatzpunkte für eine Testtheorie erschließen und umgekehrt.

Wenn hier jedoch allein jene Bedingungen analysiert werden, welche 'auf Seiten der Person' der Testreaktion zugrundeliegen, so darf nicht gefolgert werden, daß zugunsten dieser Bedingungen die determinierende Rolle der Test-stimuli nicht berücksichtigt zu werden brauchte. Bestimmte Determinanten auf Seiten der Persönlichkeit führen nicht ungeachtet verschiedenen stimulus-Materials zu gleichartigen Reaktionen. Dasselbe Motiv (charakterisiert durch eine bestimmte Motivstärke und ein bestimmtes Gefüge der übrigen Ebenen) kommt in verschiedenen projektiven Verfahren unterschiedlich zum Ausdruck : die stimuli verschiedener Tests lösen verschiedene 'Wahrnehmungshypothesen', verschiedene 'Erwartungen' und damit ein je verschiedenes Verhältnis von 'approach' und 'avoidance' aus. Doch werden diese 'Hypothesen' und 'Erwartungen' ihrerseits bestimmt durch das E b e n e n - G e f ü g e , welches diese bestimmte Motivvariable charakterisiert [2)]. D i e s e Rolle des Ebenen-Gefüges für die Reaktion auf e i n e n Test zu analysieren, ist Aufgabe dieser Arbeit. Es geht nicht um die Gegenüberstellung der Funktions-

---

1) vgl. hierzu die Einleitung, S. 28 und S.34ff.
2) Die Art der Wahrnehmungshypothese hängt z.B. mit davon ab, ob es sich um ein unterdrücktes oder ein verhaltensmäßig realisiertes Motiv handelt.

weise verschiedener Tests : hier wäre die unterschiedliche Rolle der stimuli von entscheidender Bedeutung. Bei der Analyse eines einzelnen Tests kann stattdessen das Hauptgewicht auf die persönlichkeits-spezifischen Determinanten des Testverhaltens gelegt werden.

Allerdings ist auch für eine solche Aufgabe noch einmal ausführlicher zu begründen, weshalb den Bedingungen des stimulus-Materials, deren Rolle in den hier zitierten Untersuchungen teilweise vorrangig war, eine derart untergeordnete Bedeutung zugesprochen wird. Ferner ist genauer zu bestimmen, in welchem Sinne das Ebenen-Konzept als theoretischer Rahmen für eine construct-Validierung dienen kann und wo seine Grenzen für eine solche Funktion liegen.

Z W E I T E R T E I L

Vorüberlegungen zu einem experimentellen Zugang zum Konzept der Funktionsebenen

III. Das Ebenen-Konzept als Rahmen einer construct-Validierung

1. Vorbemerkungen

Im vorangehenden Kapitel wurde erwogen, ob das Ebenen-Konzept, gefaßt als ein persönlichkeitstheoretisches Modell, einen entscheidenden Beitrag zur Theorie projektiver Methoden liefern könne. Es wurde die These aufgestellt, daß das Ebenen-Konzept dazu geeignet sei, die anhand verschiedener allgemein-psychologischer Modelle gewonnenen Erkenntnisse über die Bedingungen von Testreaktionen in einem einheitlichen Konzept zu integrieren, und zwar dies, indem vermittels des Ebenen-Modells ein theoretischer Zusammenhang hergestellt wird zwischen T e s t - Reaktion und jenem Verhalten, dessen Analyse oder dessen 'Verständnis' Ziel des diagnostischen Vorgehens ist.

Diese These besagt, daß mit Hilfe von Ebenen-Begriffen eine Theorie konzipiert werden könne, welche sowohl den Test-Index, wie das durch diesen indizierte Verhalten erfaßt. Die constructs dieser Theorie würden - auf der Basis empirischer Beziehungen - definieren, was durch den betreffenden Test indiziert wird.

Das Modell einer solchen Gültigkeitsbestimmung von Testverfahren liefert die construct-Validierung. Ihr Ziel einer vollständigen Test- und Verhaltenstheorie aber kann nur mit Hilfe umfassender empirischer Analysen durch eine schrittweise experimentelle Annäherung erreicht werden. Welcher Stellenwert kann also einem Ebenen-Ansatz auf solchem Wege der - wie Hörmann (1961, 1964b) es nennt - 'sukzessiven Approximation' zugesprochen werden?

Wir wollen in diesem Kapitel zunächst einige Probleme be-

handeln, die mit der Verwendung eines Ebenen-Modells für die Analyse des Testgeschehens verknüpft sind. Aus dem in Teil I dargelegten Material nämlich ergeben sich schon an dieser Stelle einige Einwände, welche den Ertrag einer Ebenen-Konzeption für eine solche schrittweise Gültigkeitsbestimmung kritisch einschränken. Erst anhand ihrer Diskussion ist dann jene theoretische Grundlage geschaffen, durch welche die Neuformulierung eines Ebenen-Konzeptes - in den beschriebenen Grenzen - als gerechtfertigt gelten kann.

Ein entscheidender Einwand besteht - wie oben bereits erwähnt - darin, daß mit dem Begriff der Ebene ausschließlich persönlichkeitsspezifische Bedingungen der Testreaktionen beschrieben werden und z.B. nicht die (in verschiedenen Experimenten nachgewiesene) determinierende Rolle der Test-stimuli erfaßt werden kann. In diesem Zusammenhang ist weiter zu prüfen, wieweit die G e n e s e des Testverhaltens in Ebenen-Begriffen erfaßbar ist; denn es ließ sich anhand einer Reihe von Experimenten zeigen, daß es die Genese der Testreaktion ist, welche entscheidende Hinweise auf das Indizierte selbst liefert.

Schließlich sind theoretische Vorklärungen nötig über das Verhältnis von Ebenen - B e g r i f f und Test - D a t u m : sie sind notwendig sowohl dafür, daß das Ebenen-Konzept theoretisch präzisiert werden kann (Kap. IV), wie dafür, daß sich angemessene Operationalisierungen herleiten lassen (Kap.V).

Zunächst sei der Frage nachgegangen, wieweit im Rahmen eines Ebenen-Konzeptes der Begriff der stimulus-relevance zu berücksichtigen ist.

2. Zum Problem fehlender Analysen der stimulus-relevance

Jene Experimente, in welchen die determinierende Rolle der Test-stimuli analysiert werden soll, arbeiten mit eigens für einen solchen Untersuchungszweck konstruierten, zumeist TAT-ähnlichen stimuli. Sie sind durch zweierlei ausgezeichnet : die stimuli sind - in verschiedener Abstufung - auf eine einzige Variable ausgerichtet (z.B. sex, aggression ..); es können die Reaktionen auf verschiedene 'stimulus-Stärken' (ver-

schiedene Grade der stimulus-relevance für eine bestimmte Variable) einander vergleichend gegenübergestellt werden.[1]

Eine Untersuchung jedoch, deren Gegenstand einer der gebräuchlichen projektiven Tests ist, hat es mit stimuli zu tun, die - ex definitione - mehrdeutig sind : und zwar 'mehrdeutig' nicht in dem engeren Sinne, daß sie auf eine bestimmte Variable in nur weniger direkter und augenfälliger Form verweisen (geringe 'stimulus-relevance' für z.B. sex), sondern daß sie die Äußerung vielfältiger Inhalte auslösen sollen. Sie sind als 'Mittel' für den Ausdruck von spezifisch individuellen Inhalten, nicht als 'Maßstab' für bestimmte Variablen gedacht. Der Proband soll nur angeregt werden, seine 'private world' (Leary) zu entfalten oder - wie Wilde es für die Wunschprobe formulierte - die 'Dimensionen' seiner 'charakterologischen Dynamik' zu projizieren.

Kennzeichnet aber eine solche stimulus-Mehrdeutigkeit das Ausgangsmaterial der experimentellen Analyse, so muß entschieden werden, ob nicht auch hier - für eine systematisch-quantitative Auswertung der Testreaktionen nach einzelnen Variablen - zunächst der Grad der stimulus-relevance für die eine oder andere Variable festgestellt werden muß. Denn die 'Bedeutung' eines Test-score's hängt - so zeigen es die entsprechenden Untersuchungen - entscheidend mit davon ab, wieweit die Testreaktion durch den stimulus determiniert wurde.[2]

Um zu entscheiden, welche Rolle einer quantitativen Analyse von mehrdeutigem stimulus-Material, wie es bei der Wunschprobe vorliegt, zukommt, ist ein Gesichtspunkt aufzugreifen, der schon an anderer Stelle eingeführt wurde. Er sei hier noch einmal ausführlich diskutiert. Denn an dieser Frage nach der Rolle von stimulus-Analysen entscheidet sich, ob das Ebenen-Konzept von vorneherein als unzureichend angesehen werden muß, da es den Begriff des stimulus theoretisch nicht explizit umfaßt.

1) vgl. vor allem die Arbeiten von Epstein und Mitarbeitern.
2) Die Bewertung der Testreaktion z.B. danach, ob es sich um 'Abwehr' oder bloße 'Auslassung' handelt (vgl.S. 18), ist - diesen Untersuchungen zufolge - unmöglich, solange der Grad der stimulus-relevance unbekannt ist.

Zunächst ist darauf zu verweisen, daß sich für die M e s s u n g der stimulus-relevance bei einem durch Ambiguität gekennzeichneten Test (Wunschprobe) die gleichen Probleme ergeben, wie sie für die Interpretation der Testantworten selbst erörtert wurden. Die (oben diskutierte) Tatsache nämlich einer nicht allgemeinen, sondern differentiellen Validität besagt, daß auf denselben stimulus - durch welchen Grad an stimulus-relevance er auch charakterisiert sein mag - unterschiedlich reagiert wird. Bei vorgegebenen (also nicht speziell neu-konstruierten) Tests aber kann die 'Relevanz' der stimuli allein durch die Reaktionen auf diese erschlossen werden; und gerade diese Reaktionen (Test-Indizes) erweisen sich - weil mehrdeutig - als unzuverlässiges Maß. Wenn einzelne Vpn mit viel Aggression reagieren, andere mit wenig, einige ohne Aggression : welcherart ist die stimulus-relevance?

Diese - scheinbar - nur verfahrenstechnische Schwierigkeit, mehrdeutiges stimulus-Material quantitativ zu analysieren, erklärt sich aus jenen theoretischen Erwägungen, welche schon bei der kritischen Beurteilung der Anwendung des approach-avoidance-Konflikt-Modells leitend waren(s.S. 75f). Der Begriff des stimulus, wie der der stimulus-relevance kann nur als ein Begriff der 'experimentellen Ökologie' (vgl.Fußnote 1), S. 75) aufgefaßt werden. Das heißt : dieser Begriff ist d e f i n i t o r i s c h zu bestimmen, und dieser Definition entsprechend ist ein bestimmter Grad von stimulus-relevance operational herzustellen, - wie es bei derartigen Experimenten auch der Fall ist. Der Grad der stimulus-relevance kann jedoch nicht aus den Testreaktionen im nachhinein e r s c h l o s s e n werden : denn der Begriff würde hier zu einem Lebensraumbegriff, indem durch ihn eine subjektive Wahrnehmungsleistung, also die subjektive Auffassung des stimulus beschrieben würde. Hier müßte für jede einzelne Reaktion getrennt die 'erlebte' stimulus-Qualität erschlossen werden : "The meaning of the stimulus is the response to the stimulus" (Klein, Krech 1951, S.9). Ein solcher Begriff der 'erlebten stimulus-Qualität' jedoch ist unvereinbar mit einem objektiven Normbegriff, wie ihn der Begriff der stimulus-

relevance darstellt. Das heißt : der Begriff der stimulus-relevance kann, soll er als Begriff der experimentellen Ökologie fungieren, nur durch ein objektives Maßsystem, nicht anhand subjektiver Reaktionen definiert werden.

Noch ein weiterer Gesichtspunkt ist hinzuzufügen. Es wäre denkbar, daß - analog zu den Implikationen einer differentiellen Validität - sich Gruppen bilden lassen, die einheitlich auf dieselben stimuli reagieren (z.B. mit direktem Motivausdruck) und auch durch eine vergleichbare Auffassung der stimuli charakterisiert sind. Eine solche Bildung in sich homogener 'Unter-Populationen' würde jedoch voraussetzen, daß bereits festgestellt ist, w i e der Test bei ihnen funktioniert : eine Analyse dessen, wie hier die stimuli 'aufgefaßt' werden, erübrigte sich bereits. Denn läßt sich überhaupt durch empirische Zuordnungen eine Gruppe konstituieren, bei welcher für das Verhältnis von Index und Indiziertem eine einheitliche Norm zugrundegelegt werden kann, so ist damit implizit (und im nachhinein!) festgestellt, daß die determinierende Rolle der stimuli eine einheitliche ist.

Daraus ergibt sich, daß eine Analyse der stimulus-Eigenarten eines vorgegebenen (projektiven) Tests nicht seiner Validierung vorausgehen kann, sondern daß im Vollzuge der differentiellen Validierung eines Tests an Gruppen mit systematisch unterschiedlicher Reaktionsweise (z.B. coping, avoidance, sensitization etc.) indirekt auch die unterschiedliche Rolle der stimuli differenziert wird. Bei einer Gruppe etwa, welche aufgrund eines starken Bedürfnisdrucks bestimmte stimuli eines Tests als Hinweisreize für nDominance 'erlebt' und aufgrund weiterer Bedingungen mit direktem Bedürfnisausdruck (expression) reagiert, läßt sich erst aufgrund dieser R e a k t i o n erschließen, in welcher Weise die Test-stimuli hier 'aufgefaßt' wurden, resp. welche 'Relevanz' sie besaßen. Eine andere Gruppe würde dieselben stimuli vielleicht als Hinweisreize für Aggression auffassen. [1)]

1) Diese Tatsache inter-individuell unterschiedlicher Reizgeneralisation zwischen Test- und extra-Test-stimuli macht besonders deutlich, daß der Begriff der 'misperception' keine Grundlage für eine Theorie der projektiven

Um Mißverständnissen vorzubeugen, sei eigens festgestellt : durch den Ansatz einer differentiellen Validierung wird nicht etwa die stimulus-Seite vernachlässigt, wohl aber wird deren teilweise einseitig betontes Gewicht relativiert. [1)]

Denn nicht die stimuli als solche sind ausschlaggebend, sondern deren Wahrnehmung durch den Pb; und eben diese Wahrnehmung ist inter-individuell nicht gleichartig. Zumindest aber sind die R e a k t i o n e n auf die stimuli inter-individuell in einem Maße verschieden, daß inter-individuell gültige Normen der stimulus-relevance diesen Reaktionen nicht entnommen werden können. Wie die Gültigkeit des Tests, ist auch die Rolle der stimuli d i f f e r e n t i e l l zu bestimmen. Das aber bedeutet zugleich : die stirulus-Analyse

---

Methoden liefern kann. Denn er setzt eine 'normale' oder 'richtige' Wahrnehmungsreaktion voraus, von welcher dann 'Fehlwahrnehmungen' im diagnostisch relevanten Sinne abwichen. Dieser Ansatz hat sich von Cattell (1951, Cattell, Wenig 1952) bis zu Boesch (1960) und Heckhausen (1963) halten können. Die 'erkenntniskritische' Auffassung der Projektion bei Heckhausen beruht unserer Auffassung nach auf einem falschen Kant-Verständnis.

1) Es wird also nicht die Konsequenz Rosenzweigs (1951) gezogen, auf die Rolle des stimulus vollends zu verzichten : "Since the stimulus is regarded as ultimately defined by the response, the former can no longer serve as an infallible determiner of the latter." (S. 216) Vielmehr ist die Rolle des stimulus, verstanden als Lebensraumbegriff, noch unter einem weiteren Gesichtspunkt zu betrachten. Der Begriff 'erlebte stimulus-Qualität' ist nämlich auch dann von Belang, wenn Vpn, die sich nachweislich hinsichtlich bestimmter Motivvariablen unterscheiden, g l e i c h a r t i g auf entsprechende stimuli reagieren. Der Test versagt hier in seiner Funktion, inter-individuell zu differenzieren. Die stimuli lösen stattdessen Antworten aus, welche nicht den Pb charakterisieren, sondern offensichtlich in den items 'angelegt' sind. Es wäre dies etwa der Fall, wenn bestimmte stimuli so stark assoziativ mit irgendwelchen Kognitionen verknüpft sind, daß sie zu Auslöserreizen für eben nur diese Kognitionen würden und somit - diagnostisch gesehen - Triviallösungen hervorbrächten. Trivial wäre es z.B., wenn bei der Wunschprobe 70% aller Antworten auf das item 'Vogel' lauteten : "weil der so frei fliegt". Es handelte sich um gelernte kognitive Schemata : der Vogel als Inbegriff des freien Fliegens bzw. das freie Fliegen als Definition des Vogels. Der Begriff der differentiellen Validität würde so um einen neuen Aspekt erweitert. Gruppen wären auch danach zu konstituieren, ob für sie bestimmte

kann nicht an einer anderen als der Versuchsgruppe selbst vorgenommen werden. [1)]

Sofern sich mit Hilfe des Ebenen-Konzeptes (für bestimmte Gruppen) systematische Beziehungen zwischen Index und Indiziertem nachweisen lassen, ergeben sich Hinweise auf die Rolle der stimuli (bei diesen Gruppen) mittelbar und müssen nicht zuvor gewonnen werden. Das bedeutet für die einleitend gestellte Frage : soll ein (projektiver) Test, für welchen eine differentielle Validität zu erwarten ist, analysiert werden, so muß das verwendete theoretische Modell nicht einen objektiven Begriff der stimulus-relevance (als Teil der experimentellen Ökologie) enthalten. Ferner kann bei einer ersten Einführung und Erkundung eines Ebenen-Konzeptes, wie sie hier vorgenommen werden soll, zu einem Zeitpunkt also, zu dem noch nicht ein vollständiges Theoriengebäude entwikkelt ist, vorerst auch auf die explizite Fassung des stimulus-Begriffs als eines Lebensraumbegriffs verzichtet werden : in den differentiellen Normen, die anhand entsprechender Untersuchungen für verschiedene Populationen zu bilden wären, wäre der Hinweis auf die Unterschiede in der 'erlebten stimulus-Qualität' implizit enthalten.

Triviallösungen charakteristisch sind. Es gälte die These zu prüfen, daß ein Test bei einer bestimmten Gruppe für ein bestimmtes Klischee, bei anderen Gruppen für andere Klischees empfindlich ist.

1) In diesem Zusammenhang ist Murstein (1961) zu kritisieren. Er schlägt vor, die stimulus-Eigenheiten durch eine möglichst große Gruppe (insbesondere von Studenten) b e u r t e i l e n zu lassen. Wird die 'stimulus-relevance' aber als Mittelwert von Urteilen definiert, wird a) fälschlich ein objektiver Normbegriff (experimentelle Ökologie) aus Lebensraumdaten gewonnen und werden b) die entscheidenden inter-individuellen Unterschiede in der Auffassung der stimuli unterschlagen. Damit wird zugleich noch ein weiterer entscheidender Umstand vernachlässigt : die Wahrnehmung z.B. des stimulus 'Tiger' (Wunschprobe) zum Zwecke seiner 'Beurteilung' kann nicht verglichen werden mit der Wahrnehmung desselben stimulus unter der Instruktion zu einer 'hypothetisch-fantasiemäßigen Identifikation' mit dem Objekt (im Testverhalten).

Bevor aber das Ebenen-Konzept, wie es hier verwendet werden soll, zusammenfassend präzisiert werden kann, sind noch weitere Vorklärungen nötig. Ebenfalls anknüpfend an die vorangehende Diskussion, insbesondere des Ansatzes von Leary (vgl. Kap.I), ist ein weiterer kritischer Punkt zu erörtern : die Frage nämlich, welcherart Definition des Ebenen-Begriffs (im Gegensatz zu Leary) vorzunehmen ist, d.h. die Frage nach dem Verhältnis von theoretischem Begriff und Test-Operation.

3. Theoretische Vorklärungen zum Verhältnis von Ebenen-Begriff und Test-Operation

In Kap.I ist am Beispiel von Leary ausführlich diskutiert worden, daß die operationale Gleichsetzung von Testverfahren und Persönlichkeitsebene, nämlich von 'responses to projective stimuli' und 'the preconscious and symbolic world' theoretisch unhaltbar, wie - demzufolge - praktisch irreführend ist.

Das Verhältnis zwischen Theorie und Testverfahren bzw. dem persönlichkeitstheoretischen Begriff der 'Ebene' und den empirischen Testdaten (resp. dem Begriff der 'Funktionsebene') ist also genauer zu bestimmen. Erst diese Vorklärungen können - in kritischer Absetzung gegen Leary - die Grundlage schaffen für eine Neuformulierung des Ebenen-Konzeptes. Zugleich präzisieren sie jenen theoretischen Rahmen, innerhalb dessen an späterer Stelle sich auch die Aussagemöglichkeiten unseres Experiments genauer bestimmen lassen.

Das Verhältnis zwischen Ebenen-Begriff und Test-Operation sei hier mit Hilfe des Begriffs der 'Repräsentanz' analysiert, wie er von Holzkamp (1964) eingeführt wurde. Dieser Begriff steht für einen Ansatz, der in Ablehnung des Operationalismus an der Satzzweiheit 'theoretischer Satz' und 'experimenteller Satz' festhält und das Verhältnis zwischen der 'theoretischen Realität' und der 'experimentellen Realität' zum entscheidenden Kriterium macht für die Bewertung der

Theorie, wie der ihrer experimentellen Realisation [1]. Mit dem Begriff der Repräsentanz wird die kritische Frage gestellt, wieweit die 'gemeinte' theoretische Realität tatsächlich durch die experimentellen Bestimmungen realisiert wird und damit zugleich, wieweit die 'experimentelle Realität' repräsentativ ist für die ihr (logisch) vorgeordnete Theorie.

Jedes Testverfahren läßt sich - es scheint mir dies eine vertretbare Rezeption des Holzkamp'schen Ansatzes zu sein - als eine experimentelle Operation, als eine 'Versuchsanordnung' verstehen (dies auch im Falle seiner diagnostischen Verwendung), welche bestimmte theoretische Annahmen 'realisiert' : so realisiert z.B. ein Intelligenztest bestimmte theoretische Annahmen über - oder eine bestimmte theoretische Verwendung für den Begriff 'Intelligenz'. Das Verhältnis von Testverfahren und Persönlichkeitstheorie kann also nun näher beschrieben werden als das Verhältnis zwischen einer 'experimentellen Handlungsanweisung' zu der ihr (logisch) vorgeordneten Theorie.

Insbesondere für die projektiven Tests ist es - in dieser Terminologie gesprochen - charakteristisch, daß hier weitgehend unbekannt ist, welche 'theoretische Realität' durch sie 'repräsentiert' wird : eben dies ist ja die Frage nach ihrer Validität. Kennzeichnend ist hier, daß die Theorie nachträglich gebildet werden muß, welche die 'experimentelle Realität' erst in ihrer 'Bedeutung' (Gültigkeit) erhellt, nämlich klarstellt, was diese überhaupt 'repräsentiert'.

Die 'theoretische Realität', für welche aufzuweisen ist, daß sie durch den Test 'repräsentiert' wird, kann nur - eben dieses impliziert der Begriff der 'Repräsentanz' - u n a b h ä n g i g vom Testverfahren definiert werden. Das bedeutet

1) Dieser Begriff bezeichnet in Anlehnung an Dingler (1923) die Auffassung, daß in einem Experiment die Realitätsbestimmungen wissenschaftlicher Allgemeinaussagen durch einen sozusagen 'künstlichen' Eingriff in die Realität (und in der Form von Jetzt-und-Hier-Aussagen) 'hergestellt', 'realisiert' werden.

für die Konzeption eines Ebenen-Modells : der Begriff der 'Ebene' ist so zu definieren, daß er sich prinzipiell auch anders als durch einen Test 'realisieren' läßt. Der Ebenen-Begriff kann diesem Ansatz nach nicht - wie bei Leary - operational auf einen Test reduziert werden. Es ist vielmehr erst experimentell zu prüfen, ob der 'theoretische Begriff' oder die durch ihn bezeichnete 'theoretische Realität' tatsächlich durch den Test 'realisiert' wird. Andernfalls ist die Theorie in Hinsicht auf diesen Test zu verwerfen, oder aber der Test ist so umzugestalten, daß er für diese besondere Theorie 'repräsentativ' ist. Es wird deutlich, daß - wie es in der Konzeption der construct-Validierung ausgesprochen ist - Test + Theorie eine untrennbare Einheit bilden.

Die 'experimentelle Realität' der Testdaten ist also so zu analysieren, daß ihr Verhältnis zu der - probeweise - konzipierten 'theoretischen Realität' bestimmbar wird. Hier liegt der Ansatzpunkt für eine Realisationshandlung : die 'theoretische Realität' ist in z u s ä t z l i c h e 'experimentelle Bestimmungen' zu überführen; das theoretische Konzept ist anhand von Hypothesen über die Relationen zwischen diesen experimentellen Daten und den Testdaten zu prüfen. Dieses Vorgehen entspricht zugleich einem Prinzip der construct-Validierung : die dem Test übergeordnete Theorie enthält Annahmen über verschiedene experimentelle Korrelate eines construct's und deren gegenseitige Relationen, sowie über die Relationen dieses constructs zu den experimentellen Korrelaten anderer constructs. Erst in diesem Netz von Relationen, erst als Ort innerhalb eines solchen 'nomological network' (Feigl 1958), ist die 'Bedeutung' des construct's bestimmt.

Wenn das Verhältnis von Test und Theorie in diesem Sinne als das der 'Repräsentanz' aufgefaßt wird, ist es möglich, die 'Funktionsebene' eines Tests als experimentelles Korrelat dem Begriff der 'Persönlichkeitsebene' zuzuordnen. Werden also z.B. die theoretischen Bestimmungen der 'level of values' in einem Test experimentell 'realisiert', so kann dieser Zusammenhang dahingehend bezeichnet werden, daß dieser Test auf der 'Ebene' der 'values' funktioniere oder daß er die Persönlichkeitsebene der 'values' repräsentiere.

Da jeder 'experimentellen Realität' "unbegrenzt viele 'theoretische' Interpretationen zugeordnet werden können" (Holzkamp 1964, S.237), ist auch die Art der Theorie prinzipiell beliebig. Derselbe Test kann - u.U. - sowohl allgemeinpsychologische Begriffe einer Motivationstheorie 'realisieren' (vgl. z.B. die Arbeiten von Atkinson, McClelland), wie er als experimentelles Korrelat für Konzepte einer Persönlichkeitstheorie fungieren kann. Beide Theorien unterscheiden sich u.U. nach der 'Breite' der erfaßten theoretischen Realität und wesentlich nach den ihnen eigenen Konstruktionsprinzipien : die Entscheidung für die eine oder andere Theorie aber richtet sich - zunächst - nach der Art der wissenschaftlichen Fragestellung und ist in diesem Sinne pragmatisch [1].

Testverfahren und insbesondere die durchgehend als 'Persönlichkeitstests' verstandenen projektiven Methoden sind - pragmatisch gesehen - aber gerade dann 'wertvoll', wenn sie Information erbringen, auf welcher eine 'Persönlichkeits-Diagnose' fußen kann. Diese Information ist also in einem persönlichkeitstheoretischen Begriffssystem zu formulieren (vgl. S. 82 , Fußnote 1)). Eine solche Persönlichkeitstheorie, welche die durch den Test repräsentierte 'theoretische Realität' formulierte, wäre ihrerseits strikt an den Möglichkeiten der Diagnostik orientiert : denn sie würde nicht mehr beinhalten, als was sich diagnostisch erfassen läßt. In diesem (und nur in diesem) Sinne würde der Test definieren,was er prüft : das Definiendum nämlich wäre hier als 'theoretische Realität' bestimmt, die prinzipiell unabhängig vom Test gedacht werden kann; es wäre nicht 'operational' reduziert auf das Testverfahren als einzigem Definiens (vgl. dagegen Hörmann 1964b).[2]

---

1) Auf das entscheidendere Problem der Beurteilung des 'Wertes' einer Theorie, etwa nach dem Konzept der Hypothesenwahrscheinlichkeit (Carnap 1959), der Bewährung, dem 'degree of corroboration' (Popper 1959) oder der 'Grundbelastetheit' (Holzkamp 1964) sei hier nicht näher eingegangen.

2) Das Testverfahren könnte sich gleichwohl als die 'beste' (z.B. 'ökonomischste') Operation bewähren, den fraglichen Gegenstand zu 'realisieren'.

Allerdings kann das Ebenen-Konzept nur als ein sehr 'grober' theoretischer Rahmen zur Erfassung der Funktionsweise projektiver Tests angesehen werden, da es - für sich genommen - die ihm hier zugedachte Aufgabe, die einem Test korrespondierende 'theoretische Realität' zu formulieren, nur unvollständig erfüllt. Auf diese Grenzen wird nun einzugehen sein.

4. Grenzen des Ebenen-Konzeptes

Eine entscheidende Grenze des Ebenen-Konzeptes liegt darin, daß es nicht Begriffe liefert, welche zu einer vollständigen Beschreibung der G e n e s e des Testverhaltens geeignet sind. Daß es notwendig ist, die Genese des Testverhaltens zu erfassen, aber ging - speziell für die projektiven Verfahren [1] - aus der Durchsicht verschiedener Experimente deutlich hervor (s.Kap.II)

Begründet ist diese Notwendigkeit vor allem darin, daß die Gültigkeit eines projektiven Verfahrens nur differentiell bestimmt werden kann. Die Kriterien nämlich, welche diese verschiedenen Arten der Testreaktionen differenzieren sollen, können nicht ohne den Rekurs auf eben diese Reaktionen selbst festgestellt werden. Es sind also hier die 'Zwischenglieder' in Betracht zu ziehen, welche den korrelativen Zusammenhang von Index und Indiziertem theoretisch fundieren. In der experimentellen Literatur werden sie anhand allgemeinpsychologischer Modelle beschrieben (vgl.Kap.II).

Im Ebenen-Konzept jedoch werden diese 'Zwischenglieder' nicht begrifflich mit erfaßt. Es formuliert zwar den Tatbestand, daß in einem bestimmten Testverhalten etwa die 'level of values' zum Ausdruck gelangen könnte, doch enthält das Modell für sich genommen keinen Hinweis darauf, wie das Zustandekommen einer solchen Reaktion im einzelnen zu denken ist. Gleichwohl wurde hier die These aufgestellt (s.S. 86),

---

1) In diesem Zusammenhang lassen sich proj. Tests gemeinsam als solche Verfahren kennzeichnen, bei denen der Pb auf relativ unstrukturiertes Material zu reagieren hätte (vgl.S. 13 und S.19).

daß es mit Hilfe des Ebenen-Konzeptes möglich sei, die Bedingungen (auf Seiten des Subjekts) zu formulieren, unter welchen mit der Repräsentanz der einen oder anderen 'Ebene' im Test zu rechnen sei. Welches ist also das Verhältnis zwischen dem Begriff der 'Ebene' und jenem der Genese des Testverhaltens? Die Argumentation ist hier ähnlich der des stimulus-Begriffs. Sie knüpft insbesondere an jene Diskussion an, in welcher allgemeinpsychologische und persönlichkeitspsychologische Begriffsbildung einander gegenübergestellt wurden (vgl.S.82ff).

Wir hatten dort zwei verschiedene, jedoch nicht voneinander zu trennende Aufgaben herausgestellt : die Aufgabe, den Test-Index als Ausdruck einer R e a k t i o n zu analysieren und die Aufgabe, den Test-Index als H i n w e i s auf eine diagnostisch relevante Tatsache (das Indizierte) zu erfassen. Die Begriffe des Ebenen-Modells wurden dabei benutzt, letztere Aufgabe zu erfüllen, - und zwar dies auf der Basis der Ergebnisse der ersteren. Denn die Frage nach dem Indizierten richtet sich nicht auf das a k t u e l l e Zustandekommen (die Genese) des Test-Index' (doch setzt es dessen Kenntnis voraus), sondern auf jene Bedingungen der Reaktion, welche auch über die Testsituation hinaus als 'Konstanten' (verschiedener zeitlicher oder ökologischer Erstreckung, vgl. Kaminski 1959) für die betreffende Person charakteristisch sind.

Dem Ebenen-Konzept kommt also nur eine andere 'Ordnungsfunktion' zu; es bleibt auf die Grundlage allgemeinpsychologischer Modelle 'angewiesen'. Anhand dieser Modelle werden sozusagen die einzelnen 'Steine' gewonnen, welche mit Hilfe eines persönlichkeitstheoretisch orientierten Konzeptes dann zu einem 'Mosaik' verarbeitet werden. Den 'Plan' zu diesem 'Mosaik' bestimmen die Erfordernisse der Persönlichkeits-d i a g n o s t i k . Auf dieses notwendig wechselseitige Verhältnis von allgemeiner Psychologie und Persönlichkeitspsychologie hat etwa Allport (1959) überzeugend hingewiesen. Eine vollständige Integration beider 'Ordnungsgesichtspunkte' dürfte jedoch einem noch ferneren Zeitpunkt vorbehalten sein (vgl.S.82 , Fußnote 1)).

Noch einem weiteren möglichen Einwand ist zu begegnen. Aus dieser Gegenüberstellung zweier verschiedener 'Aufgaben' oder zweier verschiedener 'Ordnungsgesichtspunkte' geht nicht hervor, in welcherart die gesuchten T r e n n k r i t e r i e n zu formulieren sind. Eben diese sollen ja anhand des Ebenen-Konzeptes festgestellt werden (vgl.S.89). Warum nicht in allgemeinpsychologischen Begriffen?

Hier ist Folgendes zu vermerken : jene Tendenzen, welche in den allgemeinpsychologischen Modellen beschrieben werden (etwa coping oder avoiding, facilitation oder inhibition, sensitization oder defense, revealing oder conceiling),sind - so kann man den entsprechenden Untersuchungen entnehmen - nicht intra-individuell konsistent [1]. Sie sind auch bei e i n e m Individuum nach Motiv-Variablen getrennt zu betrachten und liefern damit keine allgemeinen Differenzkriterien. Das Ebenen-Modell nun ist gerade dadurch ausgezeichnet, daß es jeweils auf e i n e Variable gerichtet ist (vgl.S. 85f). Indem es die Ausprägung einer Variablen auf verschiedenen 'Ebenen' beschreibt, formuliert es in diagnostisch verwertbaren Begriffen jene Bedingungen, welche obigen allgemeinpsychologischen gefaßten Tendenzen zugrundeliegen. Es erscheint daher als ein besonders geeignetes Instrument, Trennkriterien für unterschiedliche Testreaktionen zu gewinnen. Eben dieses gilt es zu erkunden.

Eine entscheidende Grenze des Ebenen-Konzeptes für den Zweck einer construct-Validierung darf jedoch nicht übersehen werden : es setzt nämlich voraus, daß man sich darüber klar sei, was ein 'Motiv', was eine 'Eigenschaft' ist. Denn anders könnte man die 'Erscheinung' dieser Eigenschaft nicht auf den verschiedenen 'Ebenen' vergleichend betrachten. Das Ebenen-Konzept enthält keine andere Möglichkeit, als die Variablen allein auf einem phänomen-analytischen Wege zu bestimmen und sie allein auf der Grundlage solcher inhaltlichen Definition

---

1) vgl. u.a. Abrams 1962, Zimbardo 1964, Sherwood 1966.

auf den verschiedenen 'Ebenen' zu untersuchen. [1] Es liefert keine Theorie etwa der 'Aggressivität', sondern kann nur ausgehen von einer definitorischen Setzung. Es liefert daher auch keinen Beitrag zur construct-Validität eines Tests i.b. auf eine einzelne Variable (verstanden als construct), also etwa 'Aggressivität', denn eine solche setzte eine Theorie der Aggression voraus (vgl. dazu etwa das Beispiel bei Hörmann 1964b, S. 37).

Damit zusammenhängend, ist die Verwendung des Ebenen-Konzeptes als Rahmen für einen Erkundungsversuch ferner dahingehend begrenzt, daß die Ergebnisse nicht über die besondere experimentelle Anordnung hinaus verallgemeinert werden können : da nämlich eine standardisierte Auswertungsanweisung u.a. für die Wunschprobe nicht vorliegt (welche nämlich eine allgemeinverbindliche Definition der 'Eigenschaften' voraussetzte), sind die Ergebnisse beschränkt auf die Einheit von Test + Theorie + Auswertungssystem.

Eine weitere Begrenzung liegt darin, daß das Ebenen-Konzept nicht den Schluß auf 'tatsächliches Verhalten' zuläßt : es enthält weder Aussagen über die Vielfalt möglicher Verhaltensmanifestationen (etwa von 'Angst'), noch wird die Auswirkung von Situations-Determinanten theoretisch mit umfaßt. Allerdings läßt es durchaus eine B e s c h r e i b u n g von offenem Verhalten - nach Maßgabe der verwendeten Definitionen - zu. Darin steht es jenen Arbeiten nicht nach, die in Kap.II referiert wurden : auch dort werden die Eigenschaftsbegriffe vorwiegend alltagssprachlich verwendet, sowohl zur Beurteilung des ('durchschnittlichen') offenen

1) Dieses ist mit ein Grund dafür, daß für die Konzeption der im Experiment verwendeten Variablen das Murray'sche need-System herangezogen wurde (Murray 1939). Murray gibt eine phänomenanalytische Beschreibung bzw. Definition seiner Variablen und leitet auf dieser Basis ab, wie diese Variablen sich inhaltlich im TAT manifestieren. Der manifeste sprachliche Gehalt von Testantworten läßt sich also anhand solcher Definitionen nach einzelnen Motiv-Variablen differenzieren. Auf derselben Grundlage beruhen - implizit oder explizit - alle in Kap. II referierten Arbeiten.

Verhaltens (bei den soziometrischen Verfahren), wie bei der Interpretation des manifesten Gehalts der Testantworten. Die Eigenschaftsbegriffe werden hier wie dort nicht als 'constructs' im strengen Sinne theoretisch fundiert : hier wie dort wird die 'Eigenschaft' als Disposition vorausgesetzt, Fragen der Generalität oder Konsistenz, der zeitlichen oder ökologischen Erstreckung werden vernachlässigt. Hier wie dort wird mit einem Durchschnittsmaß gearbeitet (siehe dazu später). In diesem Punkt geht unsere Arbeit also nicht über die zitierten Untersuchungen zu projektiven Tests hinaus.

IV. Neufassung des Ebenen-Konzepts

1. Das Ebenen-Konzept als Beitrag zur Test-Theorie

In der bisherigen Auseinandersetzung mit den vorliegenden Ebenen-Konzepten, sowohl wie mit der experimentellen Literatur zu den projektiven Methoden ist der hier zu konzipierende Ansatz bereits implizit entwickelt worden. Er sei noch einmal zusammenfassend dargestellt.

Ausgangspunkt für diesen Ansatz ist die Feststellung, daß Motive in den projektiven Methoden nicht 'direkt' erfaßt bzw. 'gemessen' werden können. Denn dem direkten Motivausdruck stehen hemmende Tendenzen entgegen, so daß er gehindert bzw. modifiziert werden kann. Weiter wird angenommen, daß ein und derselbe Test auch Reaktionen auf anderen 'Ebenen' als derjenigen der Motive hervorrufen könne. Erst die gleichzeitige Betrachtung von 'approach und avoidance', von 'need und regulation' ermöglicht - so war aufgezeigt worden - einen angemessenen Zugang zum Test-Index, nämlich zu seiner Genese und damit zum Indizierten selbst.

Entscheidend für diese Auffassung zweier gegenläufiger Tendenzen ist, daß sie als wirksam sowohl in der Testsituation, als dem M e d i u m der Diagnose, wie im alltäglichen Verhalten, als dem G e g e n s t a n d der Diagnose, angenommen werden. Nicht nur die Testreaktion wird demnach als Resultante zweier entgegengesetzt gerichteter 'Kräfte' angesehen, sondern das (alltägliche) Verhalten selbst : der unmittelbaren Bedürfnisbefriedigung stehen regulative Kräfte entgegen, welche sich auf das Motiv hindernd, modifizierend, aufschiebend oder auch verstärkend auswirken können [1].

Die Art der Bedürfnisbefriedigung, wie der Bedürfniskontrolle auf Seiten des (alltäglichen) Verhaltens wurde als entscheidend angesehen für die Art des Motivausdrucks im projektiven Verfahren : ein unbefriedigtes, weil stark kontrolliertes Motiv wird zu einer andersartigen Testreaktion

1) Auf das Konzept der 'Umwandlung' von Motiven, ihrer 'Sublimierung', sei hier nicht eingegangen.

führen als ein befriedigtes, weil verhaltensmäßig realisiertes Motiv, und dieses wird im Test anders repräsentiert werden als ein konfliktbesetztes Motiv, dessen tatsächlicher Realisation (overt behavior) kognitive Ablehnung gegenübersteht, wo also die 'Kontrolle' versagt.

Es werden hier jedoch nicht die approach- und avoidance - Tendenzen in der Testsituation jenen von Antrieb und Kontrolle im Alltagsverhalten gleichgesetzt, wie es weitgehend bei der Anwendung des Konfliktmodells zur Beschreibung des Testverhaltens der Fall ist ('tatsächliche' Aggressions-Angst = avoidance). Es wird hier vielmehr mit Epstein (vgl. S.68ff) unterschieden zwischen der 'avoidance' von Motivausdruck im Test (inhibition) und der Kontrolle eben desselben Motivs im Verhalten : ein im Verhalten realisiertes (also nicht 'verhindertes') Motiv kann in der Testantwort umgangen werden (z.B. sex); ein an seiner verhaltensmäßigen Realisation verhindertes Motiv kann im Test (substitutiv) zum Ausdruck gelangen (z.B. aggression). Diese Unterscheidung von expression-inhibition auf der einen Seite (Testreaktion) und Antrieb-Kontrolle auf der anderen Seite (im alltäglichen Verhalten) verweist jedoch nicht - das sei noch einmal betont - auf eine Unabhängigkeit zwischen den Reaktionstendenzen auf der einen und der anderen Seite, sondern verweist stattdessen um so schärfer darauf, daß eine gegebene Testantwort nicht als solche zu werten, sondern präzise auf ihre Bedingungen hin zu untersuchen ist.

Die vorstehend referierten Experimente haben deutlich gemacht, daß es nicht nur Motive sind, welche in direkter oder durch Kontrolle modifizierter Form in projektiven Tests zum Ausdruck gelangen. Auch deren 'Kontrolle' selbst, nämlich der Anspruch, sich 'realitätsangemessen' zu verhalten ('drive-reducing inhibition', vgl. S. 69), kann widergespiegelt werden. Andere Untersuchungen haben nachgewiesen, daß - unter bestimmten Umständen - ein projektives Verfahren auch als Indikator für das 'overt behavior' (gemessen anhand soziometrischer Einschätzung)

dienen kann (z.B. Mussen, Naylor 1954; Lesser 1958,1959b).

Die Bedeutung, welche demnach den Konzepten Antrieb, Kontrolle und offenes Verhalten für die Analyse des Verhältnisses von Index und Indiziertem zukommt, läßt es als gerechtfertigt erscheinen, eben diese Begriffe zur Grundlage eines Ebenen-Modells zu machen [1]. In Ebenen-Begriffe übersetzt (also weniger als 'Tendenz' denn als 'Bereich' aufgefaßt), läßt sich die 'Antriebsseite' als die 'Ebene der Motive', die 'Kontrolle' als die 'Ebene des Ich-Ideals' bestimmen.

Wir übernehmen hier die sehr einfache Klein'sche Definition des need-Begriffs als 'a want of something', welcher bestimmt ist durch die hypothetischen Merkmale von 'intensity' (tension), 'direction' und 'modifiability of tension' (1954). In dieser Definition aber ist schon die grundsätzliche Schwierigkeit enthalten, daß der Begriff 'want' nur als 'erlebtes Bedürfnis', die zusätzlichen Merkmale (intensity etc.) aber als Definitionsbestimmungen eines 'hypothetical construct's' zu fassen sind. Wie dieses Problem hier zu lösen ist, sei erst bei der Besprechung der experimentellen Vorgehensweise diskutiert (s.S.140ff). [2]

Das 'Ich-Ideal' ist definiert als Inbegriff der 'introjizierten sozialen Normen', soweit sie subjektiv als 'Anspruch an das eigene Verhalten' erlebt werden. Als dieser 'Verhaltensanspruch', (welcher nicht immer 'aktuell' erlebt werden muß, aber prinzipiell dem Bewußtsein zugänglich ist) kommt dem Ich-Ideal eine kontrollierende Funktion gegenüber der unmittelbaren Bedürfnisbefriedigung zu. [3]

---

1) Das Ebenen-Modell dient ja nicht einer Theorie der Persönlichkeit im weitesten Sinne, sondern einer diagnostisch verwendbaren Theorie projektiver Methoden.
2) Es sei ferner betont, daß es nicht Aufgabe dieser Arbeit ist, eine vollständige Motivationstheorie zu entwickeln.
3) Auf die Fülle soziologischer und psychologischer Theorien, in welchen dieser Begriff eine Rolle spielt, kann hier nicht eingegangen werden.

Ein Begriff, der hier nicht (oder besser 'noch' nicht) in die Konzeption aufgenommen ist, ist der der 'Realitätsanpassung'. Zwar gehen Realitätsbestimmungen durchaus teilweise in das Ich-Ideal mit ein [1]. Doch ist die Rolle situativer Determinanten getrennt von den kognitiven Wertvorstellungen zu fassen, da sich der aktuelle Umwelt-Druck unabhängig vom Verhaltensanspruch (diesem u.U. sogar entgegengesetzt) auf das 'tatsächliche' Verhalten auswirkt oder zumindest auswirken kann. 'Verhalten' wäre also präziser als Resultante aus Antrieb, Ich-Ideal und Realitätsbedingungen zu bestimmen.

Daß diese situativen Determinanten hier unberücksichtigt bleiben, hat zweierlei Gründe. Die situativen Bedingungen des T e s t verhaltens bestehen in den Test-stimuli und in der Testsituation (einschließlich Versuchsleiter), wie sie vom Pb erlebt wird. Auf welchem Wege die Rolle der Teststimuli bestimmt werden kann, wurde bei der Diskussion des Begriffs 'stimulus relevance' erörtert. Die übrigen situativen Determinanten des Testverhaltens zu untersuchen, wäre eine eigene experimentelle Fragestellung und nicht die dieser Arbeit. Der zweite Grund besteht darin, daß es bis heute keine Testverfahren gibt, welche ihrerseits die situativen Bedingungen des o f f e n e n (tatsächlichen) Verhaltens berücksichtigen. Sie lassen (wenn überhaupt) nur Aussagen zu über 'Verhaltensbereitschaften', nicht aber über 'tatsächliches Verhalten in bestimmten Situationen'.
Auch der Begriff des 'offenen Verhaltens', soweit er als 'Durchschnittsmaß' realisiert wird, bezeichnet eine (von außen) feststellbare Verhaltens- t e n d e n z , nicht 'aktuelles' Verhalten. Insofern es also bei der Analyse der Aussagemöglichkeiten projektiver Tests um in diesem Sinne allgemeine, und nicht nach Situationsbestimmungen gesonderte Relationen geht, genügt es, das Begriffspaar von Antrieb und Kontrolle durch die 3 Ebenen 'Motiv' (Wunsch), 'Ich-Ideal' und 'Verhalten' zu repräsentieren.

---

1) derart: 'in der und der Situation sollte man sich so und so verhalten'.

Das Ebenen-Modell erfaßt jedoch nicht nur - das dürfte aus der bisherigen Argumentation hinreichend deutlich geworden sein - die 'Bereiche' von Ich-Ideal, Motiven und offenem Verhalten als solche; es erfaßt zugleich auch das gegenseitige Verhältnis dieser 'Bereiche'. Ein solches Modell der gegenseitigen Relationen verschiedener 'Ebenen' entworfen zu haben, ist das entscheidende Verdienst von Leary (s.S. 42ff). Eine solche Theorie steht und fällt jedoch damit, ob sich ein Meßinstrument finden läßt, anhand dessen die verschiedenen 'Ebenen' metrisch miteinander verglichen werden können. [1] Eben dieses gilt auch für das hier entwickelte Ebenen-Modell. Es kann nur dann einen Beitrag zur Testtheorie liefern, wenn das Verhältnis zwischen Motiv, Ich-Ideal und offenem Verhalten sich experimentell realisieren und damit quantitativ bestimmen läßt.

Im Gegensatz zu Leary wird hier jedoch nicht mit operationalen Definitionen gearbeitet, sondern es wird - wie bereits oben ausgeführt - an der Satzzweiheit von 'theoretischem Satz' und 'experimentellem Satz' festgehalten. Daraus folgt für die experimentelle Realisation, daß sie danach bewertet werden kann und muß, wieweit sie die 'theoretische Realität' tatsächlich 'repräsentiert'. Die Operation definiert also hier nicht die Begriffe der Theorie, sondern kann sich sogar als unangemessen für die Theorie erweisen. [2] In diesem und nur in diesem Sinne kommt indessen der Operation eine entscheidende Rolle zu, denn sie bestimmt jene Grenzen, innerhalb derer die - vermittels dieser Operation gewonnenen - Ergebnisse generalisiert werden können. Die Ergebnisse sind es also, welche durch die Operation bestimmt und nur in Relation zu dieser zu beurteilen sind; nicht sind es die theoretischen Begriffe,

---

1) Auf das Meß-Instrumentarium von Leary sei nicht eingegangen, da es sich zur Analyse projektiver Verfahren (wie ausgeführt) nicht eignet.

2) Die Operation drückt zwar die hier vorgenommene 'experimentelle' Definition aus, doch kann sich die experimentelle Definition als unzulänglich für den theoretischen Begriff erweisen. Die 'Brauchbarkeit' der durchgeführten Operation wird also nicht vorausgesetzt, sondern wird im experimentellen Teil zu beurteilen sein.

welche durch eine derartige Handlungsanweisung vollständig definiert würden. [1]

Dementsprechend wird auch der theoretische Begriff der 'Differenz (Übereinstimmung) zwischen 2 Ebenen' als prinzipiell unabhängig von der Art seiner experimentellen Realisation verstanden. Es wird ausgegangen von dem 'Phänomen', daß Vpn sich inter-individuell nach dem Grad ihrer 'Motivstärke' oder etwa nach dem Grad ihrer 'verhaltensmäßigen Aggressivität' unterscheiden; und entsprechend von dem weiteren 'Phänomen', daß es - beispielsweise - unterschiedliche Anspruchsniveau-Setzungen gibt und demzufolge unterschiedlich hohe Differenzen zwischen Anspruchsniveau und tatsächlicher Leistung; oder auch von dem Faktum, daß inter-individuell verschiedenartige Differenzen zwischen der introjizierten Norm, inaggressiv zu sein und dem Grad tatsächlicher Aggressivität bestehen. Eine experimentelle Anordnung oder Operation ist dann dieser 'theoretischen Realität' angemessen, wenn sie derartige inter-individuelle Unterschiede hinsichtlich der 'Differenz zwischen 2 Ebenen' hervorbringt, realisiert.

Lassen sich die Begriffe Motiv, Ich-Ideal und offenes Verhalten [2] darüber hinaus auch u n a b h ä n g i g von dem in Frage stehenden Test (hier der WP) realisieren, so ist damit nun ein Zugang zum Problem der 'Funktionsebene' des Tests selbst gegeben. Es kann so durch Interkorrelation der Variablen festgestellt werden, ob ein Test (der ja hier verstanden wird als 'experimentelle Realität') den Begriff des Ich-Ideals, der Motive oder des offenen Verhaltens (vgl. Fußnote 2)) 'repräsentiert'. Freilich läßt sich

---

1) vgl. hierzu die Bewertung des Operationismus nicht als eine 'Theorie', sondern als 'Fußnote' zu dem allgemeinen wissenschaftstheoretischen Problem der "Definierbarkeit unseres deskriptiven Vokabulars mit Hilfe eines Teils desselben" (Bergmann, 1954).

2) Dieser Begriff wird weiter unten noch zu modifizieren sein, da er hier nicht experimentell realisiert wird.

diese 'Repräsentation' wiederum nur so weit nachweisen, wie diese Konzepte ihrerseits im Kriteriumsverfahren experimentell realisiert wurden. Eine solche zureichende Realisation einmal vorausgesetzt, läßt sich nun die Behauptung prüfen, daß bei einem projektiven Test jeweils zu unterscheiden und eigens festzustellen sei, auf welcher 'Ebene' es funktioniert hat. Es könnte so der Versuch, projektive Verfahren und 'private world' auf dem Wege operationaler Definition einander gleichzusetzen (Leary), empirisch begründet zurückgewiesen und dem Ansatz von Rosenzweig eine experimentelle Grundlage gegeben werden (vgl.S.36).

Bevor nun die Möglichkeiten einer experimentellen Realisation dieses Ansatzes im einzelnen zu prüfen sind, sei das Verhältnis dieses Persönlichkeitsmodells zu anderen Persönlichkeitstheorien (einschließlich Leary und Rosenzweig) näher bestimmt.

2. Das Ebenen-Konzept, verglichen mit anderen Persönlichkeitsmodellen

Es ist zunächst noch einmal auf die Auffassung des Ebenen-Begriffs bei Rosenzweig und Leary einzugehen. Denn das hier entwickelte Ebenen-Konzept soll ja den Ansatz insbesondere dieser Autoren präzisieren. Der hier vertretenen Auffassung nach ist das Ebenen-Modell an den Möglichkeiten und Grenzen der Diagnostik (also der projektiven Diagnostik) auszurichten; und das heißt : die Definition der 'Ebenen" ist zu orientieren am Maßstab der 'diagnostischen Relevanz' und damit an jenen 'Aussagemöglichkeiten' projektiver Tests, welche für die Persönlichkeitsdiagnose von primärem Interesse sind.

Daraus folgt etwa, daß nicht allen nur erdenklichen Meinungen, die auf dem Wege von 'opinion behavior' in einem Test geäußert werden können, ein systematischer Ort in dem hier konzipierten Ebenen-Modell zugewiesen wird, wie ebenfalls nicht alle nur erdenklichen Inhalte der 'private world' (nach Rosenzweig alle - latenten – Inhalte, die an

ihrem direkten Ausdruck willentlich gehindert werden) in dieses Konzept aufgenommen werden. Nicht alle Meinungen sind von allgemeinem diagnostischen Interesse (etwa bestimmte Stereotype oder Klischees) : eine hervorragende Rolle aber spielt das Selbstbild der Person (in etwa der level II der conscious communication bei Leary entsprechend,S.44f)[1] und jene Meinungen, an denen die Verhaltenskontrolle orientiert ist (Ich-Ideal).

Sind demgegenüber zwar u.U. alle Inhalte der 'private world' (Leary), so weit sie implizit geäußert werden, von diagnostischem Belang, so können jedoch nicht alle derart vor dem Zugriff der Umwelt (resp. des Diagnostikers) gehüteten Inhalte experimentell analysiert und aufgedeckt werden. Eine Einschränkung dieser Vielfalt denkbarer Inhalte ist daher notwendig. Die Einschränkung insbesondere auf die 'Ebenen der Motive' ist gerechtfertigt, da - das zeigen die einschlägigen Experimente - gerade der Motivausdruck im projektiven Verfahren systematisch analysiert und interindividuell verglichen werden kann. Ferner ist eine solche Einschränkung des Begriffs der 'private world' auf die 'Motive' als relativ konsistenten Determinanten des Verhaltens nahegelegt auch dadurch, daß ihre Feststellung diagnostisch (wenn auch vielleicht nicht therapeutisch) wertvoller ist, als beispielsweise die von sorgsam gehüteten Erlebnissen (als einmalig und passiv 'Erlittenem'). Bei diesen Motiven nun sind nicht nur jene von diagnostischem Interesse, welche 'öffentlich' oder auf 'direktem' Wege nicht eingestanden würden, [2] sondern auch solche, welche durchaus offen zutage treten. Die Ebene der Motive läßt sich also nicht vollständig dem Begriff 'implicit behavior' unterordnen (vgl.S. 35).

Die Ebene des Ich-Ideals (der subjektiven Verhaltensnormen)

---

1) Es spielt z.B. eine entscheidende Rolle in der Schule von Rogers (1951).
2) Daß es sich um derart 'verborgene' Inhalte handelt, war ja eine Definitionsbestimmung sowohl des Begriffs des 'implicit behavior' (Rosenzweig) wie des Begriffs der 'private world' (Leary).

entspricht zwar der bei Leary gesonderten 'level of values', wird hier aber - nach der Art ihrer Erhebung - als 'conscious communication' (Leary) oder als 'opinion behavior' (Rosenzweig) verstanden. Diese Ebene ist demnach ebenso eine Präzisierung des vagen Begriffs der 'Meinung', wie es für das Selbstbild (als den Meinungen über das eigene Verhalten) gilt. [1)]

Sollen nun Allgemeinaussagen, wie sie der Begriff der Validität impliziert [2)], und wie sie auf idiographischem Wege nicht herleitbar sind, gewonnen werden können, so sind - je nach Test - bestimmte Variablen der Untersuchung zugrundezulegen, anhand derer ein inter-individueller Vergleich erst möglich wird. Im Gegensatz zum idiographischen Vorgehen werden also nur jene Daten aus dem Testmaterial herangezogen, welche sich für einen solchen inter-individuellen Vergleich eignen, und hier wiederum insbesondere jene Aspekte des manifesten Inhalts der Testantworten, welche sich bestimmten Persönlichkeits-Variablen (Eigenschaften) zuordnen lassen.

So gesehen, läßt sich das hier konzipierte Ebenen-Modell auch dahingehend kennzeichnen, daß es die Repräsentanz

---

1) Das 'opinion behavior' könnte noch durch weitere Varianten bestimmt werden, nämlich die Meinungen über das durchschnittliche Verhalten der 'anderen', über bestimmte Charakteristika von Rollenbeziehungen, also soziale Stereotype : doch sind solche 'Meinungen' nur im Rahmen bestimmter Fragestellungen (z.B. in der Vorurteilsforschung) von direktem Interesse, so daß sie hier nicht als eine eigene 'Ebene' konzipiert werden. Allerdings muß eine Testantwort - in kritischer Absicht - daraufhin untersucht werden, ob es sich etwa um ein bloßes Klischee handelt, das mit den Motivationen und dem Verhalten des Pb in keinerlei Beziehung steht. Es war schon darauf verwiesen worden, daß solchen Klischees wahrscheinlich eine bis heute noch nicht zu übersehende Rolle - gerade auch in einem differentiellen Sinne (vgl.S.97,1))- zukommt.

2) Das klinisch-ganzheitliche Vorgehen, nach welchem nicht die Test-Indizes (resp. scores) validiert werden, sondern die jeweils einmalige Kombination von 'Diagnostiker und Test', indem ein auf dem Test fußendes Gutachten mit einem Kriterium (z.B. therapeutisches Gutachten) verglichen wird, sei hier füglich nicht als 'Validierung' verstanden.

einer bestimmten V a r i a b l e 'auf verschiedenen Ebenen' oder verschiedene 'Erscheinungsmodi' einer Eigenschaft (vgl.S. 85f) beschreibt. Die solcherart fest umrissene Funktion des Ebenen-Modells macht zugleich den Unterschied dieses Ansatzes zu den Charakterologien mit ähnlichem Gliederungsprinzip deutlich. Zum Zwecke einer Abgrenzung sei kurz auf die Differenz zwischen diesem Modell und dem 'Schicht'-Gedanken in der Persönlichkeitstheorie eingegangen.

Entscheidend ist vor allem, daß der Schichtgedanke einem völlig anderen Interesse entspringt als jenem, das bei der Konzeption des hier vorgeschlagenen Ebenen-Modells leitend ist. Im Gegensatz zum Ebenen-Modell und seiner fest umgrenzten 'Ordnungsfunktion' beanspruchen die Schichten-Theorien, den 'ganzen Menschen', seine 'Natur' zu erfassen. Der Begriff der 'Schicht' zielt ab auf den 'Aufbau der Person' (Lersch), auf verschiedene autonome 'Teilpersönlichkeiten' (Rothacker), auf die 'Seinsbestände' oder die (das 'phänomenale Funktionieren')'fundierenden Substrate' (Wellek). Der Begriff zielt also nicht ab auf das je konkrete Verhalten und dessen aktuelle Bedingungen, sondern auf das - unveränderliche 'Wesen' des Menschen (vgl. den Begriff der 'induktiven Ontologie' bei Wellek, 1959).

Alle Schichtentheorien lassen sich auf die aristotelische Dreiteilung von soma, psyche und nous zurückführen und unterscheiden sich im wesentlichen nur durch die Art der weiteren Aufgliederung dieser Bereiche. Gemeinsam ist ihnen - wenn auch verschieden durchgeführt, so doch im Prinzip - die Unterscheidung zwischen 'Tiefen-' und 'Kortikalperson', - was den Eindruck einer Parallele zu den Begriffen 'Antrieb' und 'Kontrolle' nahelegen mag. Doch geht es hier weniger um mögliche auffindbare Parallelen i.b. auf das 'Gemeinte', als um den grundsätzlichen Unterschied in der Art der theoretischen Konzeption. Die Begriffe Antrieb und Kontrolle dienen allein der Analyse von 'Verhalten' [1],

1) vgl. dafür etwa Klein (1954): "Behavior, then, expresses both the pressures of the needs and forces that counteract, qualify, or facilitate their gratification in keeping with the coping strategy that is invoked to meet the requirements and claims of reality" (S.227).

während derjenige der Schicht die Beschreibung der 'Struktur' der Persönlichkeit leisten soll und somit eine 'Realität' bezeichnet, die prinzipiell unabhängig vom Verhalten 'besteht'.

Dies wird insbesondere deutlich bei jenen Theorien, welche sich auf die hirnphysiologischen Befunde über die Funktion von Alt-, Zwischen- und Neuhirn stützen. Rothacker (1935) z.B. konzipiert eine genetisch-organische Schichtenlehre, bei welcher gerade die Verschränkung von einer (phylogenetisch wie ontogenetisch orientierten) entwicklungs-geschichtlichen Theorie einerseits und 'verhaltensanalytisch' gewonnenen Begriffen andererseits zu einem unhaltbaren Begriffsrealismus geführt hat. Die 'konkreten Handlungen' werden entwicklungsgeschichtlich unterscheidbaren autonomen 'Regulationszentren' als den eigentlichen 'Akteuren' zugeordnet [1]. Die Schichten oder 'Teilpersönlichkeiten' sind gekennzeichnet als ein 'relativ abgeschlossenes, sich selbst regulierendes, nach ständigem Gleichgewicht strebendes Ganzes' (S.43) [2]. Eine derartige Realsetzung von Begriffen findet sich auch bei Wellek. Doch werden hier die Schichten nicht durch ihr vermeintliches organisches Korrelat 'begründet', sondern es wird der nicht minder unhaltbare Versuch gemacht, die phänomenologisch gewonnenen und 'ontologisch' verstandenen Begriffe empirisch zu 'verifizieren' (vgl. Wellek 1959) [3].

---

1) Über das Verhältnis zweier derart verschiedener Gegenstandskonstituierungen vgl. Holzkamp (1964)

2) Diese Personifizierung führt so weit, daß beispielsweise formuliert wird: "Welche Schicht ist jeweils begeistet? Welche erschüttert? Welche lacht? Welche zittert? Welche handelt? Welche denkt und dichtet?" (S.140); oder : "Zentrum, das lächelt", "Zentrum, das eine mathematische Aufgabe löst" (142).

3) Es lassen sich (wenn überhaupt) nur 'Aussagen' verifizieren, sofern sie (von bestimmten Prämissen ausgehend) nämlich in Handlungs-Ereignis-Relationen überführt werden können; nicht jedoch lassen sich Begriffe (wie etwa der des 'Charakterkerns' bei Wellek) verifizieren. Begriffe lassen sich nur definieren (auf verschiedene Weise), was den Irrtum des Begriffsrealismus notwendig ausschließt.

Diese auf dem Schichtgedanken aufbauenden Charakterologien sind insgesamt, da sie auf 'Seiendes' abzielen bzw. im wesentlichen Klassifikationsprinzipien für die 'hinter dem Phänomenalen' liegende 'Realität' aufstellen, für die praktisch-diagnostischen Probleme unbrauchbar : "Dem klinischen Psychologen fällt es schwer sich vorzustellen, wie diese Charakterologien auf Menschen angewendet werden können, die dringende Persönlichkeitsprobleme haben..." (Wyatt 1959, S.260) [1].

Anders ist es mit einer Theorie, gegen welche auch der Vorwurf des Begriffsrealismus erhoben werden kann, welche aber gleichwohl dem hier konzipierten Ansatz eher vergleichbar ist, obwohl sie nicht unbeteiligt ist am Entstehen gerade der Schichtentheorien : die Freud'sche Psychoanalyse. Diese kann (m.E.) - entgegen obigen Charakterologien - ihre klinische Bewährung durchaus ins Feld führen. Hier nämlich handelt es sich weniger um eine 'Strukturtheorie' (wie die Begriffe 'Ich', 'Es', 'Über-Ich' nahelegen mögen), als um eine dynamische Theorie des Verhaltens. Die 'hinter' dem Phänomenalen liegende 'Realität' interessiert nicht als solche, sondern nur soweit sie Erklärungsprinzipien liefert für das beobachtbare Verhalten (resp. die Krankheitssymptome) [2]. Auch gegen diese Position ist eine ausführ-

1) vgl. auch Hörmann 1964b

2) Daß es sich dabei um 'Erklärungsprinzipien' und nicht um eine 'Realität' im Wortsinn handelt, wird von Freud immer wieder betont ("...ob unsere Ich-Psychologie überhaupt darauf hinausläuft, gebräuchliche Abstraktionen....aus Begriffen in Dinge zu verwandeln, womit nicht viel gewonnen wäre"; in: 'Neue Folge der Vorlesungen zur Einführung in die Psychoanalyse' Bd XV, S.66); vgl. dazu auch Frenkel-Brunswik 1959. Gleichwohl läßt Freud's deskriptives Vokabular es kaum zu, eine klare Grenze zwischen 'Metapher' und 'Wirklichkeit' zu ziehen. Der Vorwurf des Begriffsrealismus ist jedoch auch schon deshalb hier weit weniger relevant, als mit Hilfe dieser Theorie 'aktuelle Verhaltensmanifestationen' und nicht 'transphänomenale Substrate' beschrieben werden. Zudem steht in der gegenwärtigen Freud-Rezeption einer Neu-Formulierung Freud'scher Konzepte nach heutigen Wissenschaftkriterien (welche damals noch gar nicht entwickelt waren !) wahrscheinlich nichts im Wege (vgl. dazu etwa Horowitz 1963, Rapaport 1960, Frenkel-Brunswik 1959, gegen letztere (wenig überzeugend) Eysenck 1959)

liche Abgrenzung hier nicht am Platze; es sei nur ein entscheidender Punkt herausgegriffen.

Es interessiert hier weniger die übliche Kritik an der Aufgliederung in 'Ich', 'Es' und 'Über-Ich', als eine Implikation dieser Aufgliederung : nämlich die Gegenüberstellung von 'Lustprinzip' und 'Realitätsprinzip'. Der unmittelbaren Bedürfnisbefriedigung (Antrieb) steht auch hier das Prinzip der Realitätsanpassung (Kontrolle) entgegen : "..zwischen Bedürfnis und Handlung wird der Aufschub der Denkarbeit eingeschaltet" (Bd XV, S.82). Die Bedürfnisbefriedigung wird also in Anpassung an die Realitätsbedingungen 'dirigiert', das Lustprinzip im Laufe der Entwicklung durch das Realitätsprinzip 'entthront'; das Verhalten selbst ist eine Resultante aus Antrieb (Es), Anspruch (Über-Ich) und Realitätsanpassung. Die Person wird somit nicht als ein nach außen hin abgeschlossenes System verschiedener autonomer 'Akteure' (Rothacker) verstanden. Vielmehr spielen in dieser Theorie die Anforderungen der Außenwelt eine entscheidende Rolle. Dieses spiegelt sich nicht nur im Begriff des Realitätsprinzips, als der aktuellen Verkörperung des 'Ich's' wider. Auch das 'Über-Ich', als Inbegriff der Verhaltensnormen, fungiert als Vermittlungsinstanz zwischen sozio-kultureller Umwelt und individuellem Verhalten (zwischen Philogenese und Ontogenese). Von letzterem ist jedoch unser Begriff des 'Ich-Ideals' abzugrenzen, denn diesem haften nicht jene Implikationen psychoanalytischen Theoretisierens an, nach welchen die Entstehung des Über-Ich's aus dem Ödipus-Komplex erklärt wird. In dieser Arbeit geht es nicht um eine 'Erklärung' des Über-Ich, sondern um eine Definition des Begriffs 'Ich-Ideal'.

Der Begriff der 'Ebene' ist in unserem Modell also nicht anders als ein Konstruktionsprinzip zu verstehen. Ihm haften nicht die Implikationen des Schichtgedankens an, welcher mit dem Prinzip der 'Aufschichtung' nicht nur falsche räumliche Vorstellungen [1] weckt, sondern auch den (system-

1) vgl. z.B. den kritischen Aufsatz von Heyde, 1956.

transzendenten) Wertgesichtspunkt von 'höher = besser' nahelegt. Der Ebenen-Begriff teilt weiter nicht verschiedene Implikationen der psychoanalytischen Theorie (welche im einzelnen auszuführen hier nicht der Ort ist), wenn auch das Prinzip von Antrieb und Kontrolle dieser Position entlehnt ist. Das Ebenen-Modell liefert keine Theorie im umfassenden Sinne, sondern ein Begriffssystem, welches nicht mehr als den Ausgangspunkt für eine Testtheorie liefert und in dieser Funktion - bei allen allgemeinpsychologischen Grundlagen - persönlichkeitstheoretisch, nämlich diagnostisch orientiert ist. Es liefert weder eine Theorie der Persönlichkeit, noch der Motivation, der Fantasie oder der stimulus-Apperzeption. Es hat nur die Funktion, die Mehrdeutigkeit von Testreaktionen zu analysieren und mit Hilfe von Ordnungsgesichtspunkten aufzuheben.

V. Konsequenzen für eine experimentelle Untersuchung

1. Vorbemerkungen

Das Ziel der experimentellen Untersuchung besteht - allgemein gesehen - darin, zu erkunden, ob das hier formulierte Ebenen-Konzept Ansatzpunkte liefert für die construct-Validierung eines projektiven Verfahrens. Bei der Durchsicht verschiedener experimenteller Arbeiten (Kap. II) ließ sich die Erwartung begründen, daß mit Hilfe des Ebenen-Konzeptes Bedingungen der Testreaktion auf Seiten des Subjekts begrifflich erfaßt und damit 'Trennkriterien' bestimmt werden könnten, welche Populationen unterschiedlicher Testreaktionen differenzieren. Ziel der Untersuchung im engeren Sinne ist es somit, vermittels des Ebenen-Konzeptes Kriterien aufzudecken, welche als Ausgangspunkt für eine 'differentielle' Validierung dienen können.

Allerdings geht es - das braucht kaum noch einmal betont zu werden - nicht darum, alle überhaupt nur vorkommenden Reaktionsformen bei einem projektiven Test als solche aufzudecken. Die Aufgabe ist enger gesteckt. Es geht ausschließlich um den Nachweis solcher Reaktionen, welche - vom Test aus betrachtet - sich den Begriffen der 'Funktionsebene eines Tests' im hier verstandenen Sinne subsumieren lassen : also nur um die Reaktionen, die entweder die 'Ebene' des Ich-Ideals, der Motive oder des offenen Verhaltens widerspiegeln. Die jeweilige Funktionsebene ist sodann in ihrer Relation zum 'Ebenen-Gefüge' der Person zu analysieren : es ist zu erkunden, ob sich so 'Trennkriterien' gewinnen lassen, anhand welcher Testreaktion und Kriteriumsverhalten sich einander (korrelativ) zuordnen lassen.

Es ist bereits dargelegt worden, daß der Begriff der 'Ebene' wie der des 'Ebenen-Gefüges' sich nur auf einzelne Variablen (also nicht etwa auf autonome 'Teilpersönlichkeiten') beziehen können, deren verschiedene 'Erscheinungsmodi' er begrifflich isoliert (vgl.S.116f). Er läßt sich anwenden also nur auf solche Variablen, die alltagssprachlich als 'Eigen-

schaften' bezeichnet werden : diese 'Verhaltenscharakteristika' nämlich lassen sich von der Antriebsseite, wie unter dem Aspekt der Verhaltenskontrolle bzw. des Verhaltensanspruchs betrachten.

Bei der Durchsicht entsprechender Experimente ergab sich weiter, daß die Form der Testreaktion [1] nicht losgelöst betrachtet werden kann von der jeweilig die Reaktion kennzeichnenden Variablen; daß also etwa unterschieden werden müsse, ob das in der Testreaktion angesprochene Verhalten schuld- oder angstbesetzt oder sozial erstrebenswert ist. Es sollten daher bei denselben Vpn Testantworten, die sozial erstrebenswerte 'Eigenschaften' und solche, die negativ bewertete 'Eigenschaften' beinhalten, einander gegenübergestellt werden können.

Es ist ferner ausführlich erörtert worden, daß sich die stimulus-Relevanz (hier für die Wunschprobe [2]) nicht im vorhinein feststellen ließe, daß ihre Kenntnis aber insbesondere für den Vergleich der Reaktionen auf verschiedene Tests notwendig sei. Diese unterschiedliche stimulus-Relevanz kommt indirekt in den unterschiedlichen Normen zweier verschiedener Tests zum Ausdruck. Daraus ergibt sich, daß sich indirekt Hinweise auf die Art der stimulus-Relevanz auch ermitteln lassen, indem man verschiedene Testverfahren unter folgender Frage vergleicht : wie reagiert eine Gruppe, die beispielsweise in der WP ihre (nicht-realisierten) Aggressions-Wünsche ausdrückt [3], auf den TAT, und zwar dies im Unterschied zu einer anderen Gruppe, die etwa im Sinne ihrer kognitiven Ablehnung von Aggressivität auf die WP reagierte. Bei einer solchen Gegenüberstellung ließen sich u.U. auch Hinweise auf die Rolle des jeweilig geforderten Reak-

---

1) Es sei daran erinnert, daß es wesentlich darauf ankommt, verschiedene 'Formen' der Testreaktion wie 'direkter Motivausdruck', 'defense', 'coping' etc. zu unterscheiden.
2) Dieser Test wird im folgenden abgekürzt als WP.
3) Die Art der Reaktion auf die WP wäre durch Korrelation mit dem 'Kriterium' festgestellt worden.

tionstyps (Wunschäußerung - Geschichte erfinden) aufdecken (vgl. S. 78). [1)]

Da bei dieser Untersuchung - vermittels des Ebenen-Konzeptes - Trennkriterien für eine differentielle Validierung gewonnen werden sollen, ist ein systematischer Gruppenvergleich erst im nachhinein, nämlich nur dann, wenn sich derartige Kriterien haben bestimmen lassen, möglich. Es können also nicht die Gruppen bereits nach vorliegenden Kriterien zusammengestellt werden, sondern müssen vielmehr anhand des Untersuchungsmaterials nachträglich gebildet werden. Dieses aber setzt voraus, daß die Versuchsgruppe zumindest nach allgemeinen soziologischen Kriterien relativ homogen ist (Alter, Intelligenz, Ausbildungsstand), damit inter-individuelle Differenzen in der Testreaktion vorwiegend auf Differenzen in der Ausprägung der untersuchten Variablen bzw. auf die festgestellten Unterschiede in den 'Ebenen-Relationen' zurückgeführt werden können und jedenfalls nicht andersartige und unübersehbare Faktoren die entscheidendere Rolle spielen.

Es sind also bei den nach einheitlichen Kriterien zusammengestellten Vpn Information über die Art ihres 'Ebenen-Gefüges' bei verschiedenen Variablen (Kriterium), sowie ein möglichst vielfältiges Testmaterial einzuholen. Auf der Grundlage eines solchen Materials sind nun folgende Fragen, die hier eher als Leitgesichtspunkte, denn als Hypothesen im strengen Sinne aufzufassen sind, zu klären :

a) Kann von einem Test aufgewiesen werden, daß er bei verschiedenen Gruppen von Vpn 'auf je verschiedenen Ebenen funktioniert'?

b) Gibt es bestimmte Variablen, die vorrangig (also durchgängig für alle Vpn) bei einem bestimmten Test im Sinne einer bestimmten 'Funktionsebene' zum Ausdruck gelangen?

c) Läßt sich zeigen, daß eine Untergruppe bei einer Variablen auf der einen Ebene, eine andere Untergruppe aber bei derselben Variablen auf einer anderen Ebene reagiert?

d) Lassen die Kriterien, nach denen diese Untergruppen differenziert werden, sich tatsächlich auch ihrerseits im Begriffssystem des Ebenen-Konzeptes formulieren?

---

1) Es sei schon an dieser Stelle darauf verwiesen, daß der TAT zwar aus eben diesem Grunde erhoben wurde, daß er jedoch aus zeitlichen Gründen nicht verwertet werden konnte.

e) Läßt sich nachweisen, daß eine bestimmte Vp bei einem Test je nach Variable im Sinne verschiedener Funktionsebenen, also inkonsistent reagiert?

2. Allgemeine Charakterisierung der experimentellen Vorgehensweise

a) Die Wunschprobe als experimenteller Ausgangspunkt

Einleitend ist bereits dargestellt worden, daß das Ebenen-Konzept beispielhaft an der Wunschprobe (WP) erkundet werden soll. An jener Stelle ist zugleich näher erläutert worden, wie dieses 'projektive' Verfahren von seinem Autor (zum Zeitpunkt seiner Veröffentlichung) verstanden wurde (s.S. 29). Seither ist keine weitere Arbeit zu diesem Test erschienen; der Autor selbst starb, bevor er umfangreicheres Untersuchungsmaterial zur WP publizieren konnte.

Die WP stellt also wohl die noch größte Leerstelle im Vergleich zu den sonst in Deutschland gebräuchlichen Verfahren dar; und schon aus diesem Grunde bietet es sich an, diesen Test einer näheren Analyse zu unterziehen. Es läßt sich aber zudem bei diesem Verfahren besonders anschaulich demonstrieren, daß die Annahme verschiedener 'Funktionsebenen' - wie sie für den TAT (bzw. TAT-ähnliche stimuli) anhand vorliegender experimenteller Ergebnisse indirekt erschlossen worden war - auch hier begründet ist. (vgl. das Beispiel S. 26)

In der Instruktion zur WP wird der Pb aufgefordert, er solle sich mit Hilfe seiner Fantasie in die Lage versetzen, daß er sich etwas w ü n s c h e n kann, nämlich sich wünschen kann, ob er die einzelnen vorgeführten 66 Gegenstände (z.B. 'Grashalm') selbst s e i n möchte oder nicht. Ein solcher Wunsch, etwas Bestimmtes zu 'sein' (z.B. ein 'hilfebringendes Sanitätsauto' oder ein 'unbezwingbarer Berg') kann verschiedenen 'Quellen' entstammen, die sich ihrerseits mit dem Konzept der 'Ebenen' beschreiben lassen.

Man kann etwas sein wollen, weil es sozial erstrebenswert ist, weil man es 'sein sollte' (z.B. hilfsbereit); oder man

kann fantasiemäßig einen Wunsch äußern, den man 'realiter' nicht erfüllen kann und 'idealiter' nicht anerkennt, weil er 'nicht sein sollte' (der Wunsch, aggressiv zu sein); oder man kann etwas (sozial Erstrebenswertes) zu sein wünschen, weil man nicht schafft, es zu realisieren (z.B. autonom zu sein). Schließlich ist denkbar (jedoch ebenfalls nachzuprüfen), daß ein item eine Wahrnehmungshypothese auslöst, nach welcher die mit dem betreffenden Gegenstand (item) vorstellungsmäßig verknüpften Verhaltensweisen (die Aggressivität des 'Tigers', oder das 'ständig-unter-Menschen-sein' beim 'Tanzsaal') dem eigenen Selbstbild entsprechen [1]. Ebenso kann die Reaktion, 'etwas nicht sein zu wollen' (negative Reaktion in der WP) verschiedenen Quellen entstammen : etwas nicht sein zu wollen, weil man es 'ist', aber 'nicht sein sollte' (z.B. aggressiv), oder weil 'kein Bedürfnis danach besteht' (z.B. einsam zu sein wie eine 'Kapelle'), oder weil es der Norm widerspricht (zu zerstören wie ein 'Orkan') oder auch, weil es dem eigenen Selbstbild widerspricht (man hält sich für inaggressiv und lehnt 'aggressive' items ab.

Für die Auswahl der Variablen, nach welchen - im Rahmen der hier gestellten Aufgaben - die Testprotokolle auszuwerten sein würden, sind bereits einige Kriterien genannt worden : sie müssen auf allen 3 'Ebenen' betrachtet werden können und sozial erstrebenswerte, wie sozial negativ bewertete 'Eigenschaften' umfassen. Beide Bedingungen erfüllt - wie im folgenden zu zeigen sein wird (s.S.130ff) - das Murray'sche need-System.

Zugleich müssen die Variablen aber überhaupt in den Testreaktionen 'antreffbar' sein : diese Entscheidung, welche needs vorwiegend Eingang in die WP finden, haben wir allein

---

1) Eine solche Reaktion entspräche nicht der Instruktion, 'Wünsche' äußern zu sollen; es kann indessen erwartet werden, daß ein Pb sich - trotz der Instruktion - fragte 'wer bin ich eigentlich' und nun diesem Selbstbild entsprechend auf die items reagiert. Es könnte eine solche Reaktion relativ durchgängig sein oder nur durch einzelne items ausgelöst werden oder nur bei bestimmten Variablen (z.B. exhibition oder dominance) vorliegen.

aufgrund von Erfahrung mit der WP, von Introspektion bzw. Reflektion über die stimulus-Qualitäten getroffen. Die so zusammengestellten Variablen wurden in einem Vorversuch überprüft und aufgrund des dort gewonnenen Testmaterials teilweise ausgetauscht. Endgültig aufgenommen wurden folgende Variablen : nAggression, nDominance, nAutonomy, nAffiliation, nNurturance, nExhibition und nHarmavoidance [1]. Im Hauptversuch konnten in fast allen Testprotokollen sämtliche Variablen signiert werden [2].

An anderer Stelle ist bereits dargelegt worden, daß ein zusätzliches projektives Verfahren - der TAT - zum Zwecke weiterer Information zwar erhoben, jedoch aus zeitlichen Gründen nicht ausgewertet wurde. (vgl.S.123f)

Schließlich ist noch auf ein weiteres Zusatzverfahren einzugehen. Doch soll auch dieses nicht eingehend erörtert werden. Denn es konnte - ebenfalls aus Zeitgründen - nicht vollständig ausgewertet werden und spielt daher für die Ergebnisse unseres Experiments nur eine untergeordnete Rolle.

Bei der Diskussion des approach-avoidance-Konflikt-Modells war aufgezeigt worden, daß Verteidigungsverhalten eindeutig nur herleitbar ist bei hinlänglich hoher stimulus-relevance (vgl.S.61ff, S.65f). Demgegenüber indiziert bei relativ mehrdeutigem stimulus-Material (geringe stimulus-relevance) derselbe geringe Test-score entweder avoidance oder aber eine geringe Ausprägung der betreffenden Variablen oder auch nur eine diagnostisch irrelevante 'Auslassung' (s.S.47). Es sollte daher Material gewonnen werden, welches zusätzliche Information über mögliches Vermeidensverhalten erbringen könnte.

In Anlehnung an Leary, der seine Ebene IV (level of the unexpressed) durch perceptual defense und selective recall näher

---

1) Die Variablen werden im folgenden ebenso wie bei Murray (mit Ausnahme von nHarm) abgekürzt als nAgg, nDom, nAut, nAff, nNur, nExh und avoid. Für eine genauere Bestimmung dieser Variablen siehe den folgenden Abschnitt.

2) Weshalb noch weitere und ebenfalls in der WP potentiell auftretende Variablen aufgenommen wurden, wird erst in einem späteren Zusammenhang begründbar (vgl.S.146).

zu bestimmen sucht (s.S.47f), sollte auch hier der Versuch gemacht werden, anhand von Daten über 'selektives Behalten' dieses zusätzliche Material zu gewinnen.

'Selective recall' ist ein heute schon vielfach untersuchter Gegenstandsbereich. Zum einen wird allgemein untersucht, welche Determinanten bei der Strukturierung der Gedächtnisinhalte eine Rolle spielen; zum anderen wird, analog zur social-perception-Forschung, insbesondere nach den motivationalen Bedingungen der Selektivität gefragt. So ist aufgewiesen worden, daß es für die Art des Wiedererinnerns entscheidend ist, ob das Material 'angenehm' oder 'unangenehm' ist [1] oder ob es mit der eigenen Meinung übereinstimmt oder nicht [2].

Ein Zusammenhang aber läßt sich auch herstellen zwischen Verteidigungsmechanismen und selektivem Erinnern. So versucht Erikson (1941) nachzuweisen, daß die Vpn, die vorwiegend ich-bedrohliche Inhalte vergessen, auch im Bereich der Wahrnehmung mit perceptual defense reagieren, wie auch zum Lernen emotional besetzter Wörter mehr Zeit benötigen. [3] Zu verweisen ist in diesem Zusammenhang auch auf die Arbeiten zum Behalten erledigter und unerledigter Aufgaben : es werden - entgegen dem sgn. Zeigarnik-Effekt - dann mehr erledigte Aufgaben behalten, wenn die unerledigten Aufgaben ich-bedrohlich sind [4]. Schließlich wird noch die Unterscheidung zwischen Vpn mit 'high ego strength' und 'low ego strength' vorgenommen [5] : es sind die Vpn mit low ego strength, welche vorwiegend bedrohliches Material erinnern [6].

---

1) Barrett 1938; Rapaport 1942
2) Edwards 1941; Levine, Murphy 1943; Levine, Gardner 1947; Alper, Korchin 1952
3) Diese 'constance hypothesis' ist allerdings, wie aufgezeigt, bisher nicht bestätigt (s.S.78f).
4) vgl. Zeigarnik 1927 und u.a. Lewis, Franklin 1944; Lewis 1944; Rosenzweig 1943; Erikson 1952
5) 'Ego strength' wird definiert als "the degree to which an individual's behavior is generally in keeping with the realities of the objective situation" (Erikson 1954,S.46)
6) vgl. Alper 1946; Carlson 1954; Erikson 1954

Aus derartigen Untersuchungen ziehen Alper und Korchin folgende Konsequenz : "Selective recall can serve as an individual or group projective device when material which is socially and personally relevant is used as the experimental material." (1952,S.36)

Für die Frage nach möglichem Verteidigungsverhalten in der WP läßt sich diesen Arbeiten folgender Hinweis entnehmen : Vpn, die vorrangig die abgelehnten items wiedererinnern, könnten allgemein durch 'low ego strength' charakterisiert sein und besonders sensitiv auf die negativen Aspekte der Umwelt reagieren. Vpn, die vor allem die items, mit denen sie sich positiv identifizieren, wiedererinnern, könnten durch 'high ego strength' charakterisiert sein und entsprechend weniger gerade ich-bedrohliche Inhalte aus dem stimulus-Material selegieren. Diese Hypothesen sind analog den Ergebnissen von Phares (1961, s.S. 77f), nach welchen ängstliche Vpn nicht defensiv, sondern sensitiv auf ich-bedrohliche cues reagieren.

Noch einmal umgekehrt formuliert, sind folgende Relationen denkbar: Vpn, welche mit 'sensitization' reagieren und mit den WP-items besonders häufig gerade ich-bedrohliche Situationen verknüpfen (z.B. bei 'Grashalm' : 'zertreten werden' anstelle von 'sich wiegen im Winde'), werden auch in der Erinnerung vorrangig die abgelehnten items behalten. Vpn dagegen, welche mit 'perceptual defense' [1] reagieren und ich-bedrohliche Inhalte weitgehend abwehren, neigen dazu, vorrangig positives Material (in der WP die items ihrer positiven Wahl) wiederzuerinnern. Schließlich könnte bei diesen Vpn, die vor allem die positiv beantworteten items wiedererinnern, nicht ego-defense ausschlaggebend sein; sondern sie

---

1) 'Perceptual defense', das sei in Erinnerung gerufen, wird hier dahingehend allgemein definiert, daß in den verbalen R e a k t i o n e n auf stimulus-Material, das als ich-bedrohlich erlebt werden k a n n , diese ich-bedrohlichen Inhalte nicht explizit auftreten (ungenannt bleiben). Die Frage, wie eine solche 'Vermeidung' zustandekommt, ist damit nicht entschieden (vgl.S.76, Fußnote 1)).

könnten durch 'ego strength', d.h. eine allgemein stärker positive Hinwendung zur Umwelt, charakterisiert sein.

Die experimentellen Hypothesen, die sich aus diesen Überlegungen ableiten lassen, sind erst im Rahmen der Darstellung des Experiments selbst einzuführen. Das Datenmaterial wird folgendermaßen gewonnen : im Abstand von einer Woche nach der Erhebung der WP werden die Vpn aufgefordert, möglichst viele der WP-items zu reproduzieren und in der Reihenfolge des Erinnerns zu notieren.

b) Die Variablen des Experiments (Murray's need-system)

Es sind verschiedene Merkmale, aufgrund deren das Murray'sche need-System sich für die Aufgabenstellung unseres Experiments besonders gut eignet.

Dieses Variablen-System, Kernpunkt der Murray'schen Persönlichkeitstheorie, wurde - das ist zunächst entscheidend - für d i a g n o s t i s c h e Zwecke entwickelt : es diente einer heute schon 'klassischen' diagnostischen Analyse an 50 Probanden, welche sich über einen Zeitraum von 2 1/2 Jahren erstreckte. Gewonnen wurden diese Variablen aufgrund einer umfassenden phänomenalen Analyse des Verhaltens : Gegenstand der Theorie ist also das Verhalten, nicht transphänomenale (ontologische) 'Strukturen' der Persönlichkeit. 'Verhalten' aber und seine Bedingungen sind auch der Gegenstand unseres Ebenen-Konzeptes.

Die Variablen, die Murray und Mitarbeiter für eine möglichst weitreichende Persönlichkeitsbeschreibung zusammenstellten, fungieren als (nicht endgültiges, sondern eher hypothetisches) Klassifikationsschema : das Ebenen-Konzept liefert - so war dargelegt worden - keine Theorie einzelner 'Eigenschaften', sondern richtet sich auf eben solche phänomenanalytisch gewonnenen Verhaltensvariablen, welche nicht 'Entitäten', sondern klassifikatorische Begriffe darstellen.

Die Murray'sche Theorie nun selbst, insbesondere der construct 'need', entsprechen in wesentlichen Zügen unserem

Ansatz [1]. 'Verhalten' wird verstanden als ausgelöst durch 'Kräfte' (force), und eben diese Kräfte werden beschrieben als einzelne isolierbare needs. Der Begriff 'need' ist definiert als "a force which organizes perception, apperception, intellection, conation and action in such a way as to transform in a certain direction an existing, unsatisfying situation". "Thus, it (the need,Verf.) manifests itself by leading the organism to search for and to avoid encountering or, when encountered, to attend and respond to certain kinds of press." "... it persists and gives rise to a certain course of overt behavior (or fantasy), which ... changes the initiating circumstances in such a way as to bring about an end situation which stills (appeases or satifies) the organism." (1962, S.124)

Der Begriff 'need' beschreibt also die (inneren) Bedingungen des Verhaltens, wie dieses selbst : der need manifestiert sich i m Verhalten. Wichtig ist dabei die Unterscheidung zwischen 'latent need' und 'manifest need'. Die latent needs sind in unserer Terminologie die 'nicht realisierten', die 'an ihrer Objektivation gehinderten' Motive. Sie spielen, obwohl nicht manifest, auch für Murray eine entscheidende Rolle: sie sind 'active fantasies which ... follow an 'irreal' course" (a.a.O. S.111). Manifest wird ein need im Verhalten. Doch die tatsächliche need-Stärke leitet sich nicht allein aus diesem ab, sondern "to determine the total strength of a need one should consider the amount of internally inhibited tension as well as the amount of externally exhibited activity" (a.a.O., S.145). Ein Bedürfnis wird also auch hier 'auf verschiedenen Ebenen' betrachtet.

Das Verhalten seinerseits ist den 'press' der Umwelt ausgesetzt und wird durch diese modifiziert. Dieser Begriff 'press'

---

1) Es sei hier nicht eingegangen auf den unglücklichen Versuch Murray's, seinen construct 'need' in Zusammenhang zu bringen mit hirnphysiologischen Prozessen. Dieser Versuch entspringt wohl insbesondere dem Bedürfnis, die wesentlich klinische Orientierung positivistisch zu 'ergänzen' oder zumindest, sich gegen den Vorwurf des 'Mentalismus' zu schützen. - Die Theorie sei nur so weit erläutert, wie es zur Kennzeichnung des von uns benutzten Variablen-Systems nötig ist.

wird ebenso wie der des 'need' vom 'effect' her definiert, von dem (erkennbaren) Effekt, den innere oder äußere Kräfte für das Subjekt resp. für das Verhalten haben. Verhalten ist also eine Resultante und ist diagnostisch faßbar nur vermittels der Bestimmungstücke, aus denen es resultiert : "the complete behavioral event (appears) as an interaction of forces" (a.a.O.,S.118).

Auch das 'super-ego-system' wird - dem Ebenen-Ansatz vergleichbar - als eine Determinante des Verhaltens konzipiert : es fungiert als Kontrollinstanz oder Regulativ gegenüber den needs, denen es die (verinnerlichten) kulturellen Verhaltensnormen entgegensetzt. [1)]

Die Verhaltensvariablen nun, welche sozusagen das 'Anschauungsmaterial' für dieses theoretische Begriffsgerüst liefern, bezeichnen einzelne begrifflich isolierbare needs. Diese werden ihrerseits definiert anhand ihrer beobachtbaren Manifestationen. Mit den gleichen Variablen lassen sich daher sowohl die motivationale Seite, wie deren verhaltensmäßige Objektivation beschreiben, wie auch jene Normen, denen dieses Verhalten entsprechen soll. Diese Variablen lassen sich also - in unserer Terminologie gesprochen - auf allen '3 Ebenen' betrachten.

Weiterhin ist durch die sehr ausführlichen Murray'schen Definitionen eine begriffliche Grundlage auch für die folgende Aufgabe gegeben : die Testprotokolle danach auszuwerten, welche dieser Variablen sie ihrem s p r a c h l i c h manifesten Gehalt nach enthalten (vgl. dazu S.192ff). Anhand eines solchen Begriffsgerüsts ist es möglich, die umfangreichen (weil nicht standardisierten, sondern freien) Protokolle der Wunschprobe nach einzelnen Motivkategorien inhaltlich zu formalisieren.

---

1) Damit keine Verwechslungen entstehen, sei folgendes angemerkt. Außer diesem Begriff des 'super-ego-system' verwendet Murray auch den Begriff des 'ego-ideal'. Dieser bezeichnet aber nicht, wie unser Begriff des Ich-Ideals (ideal-self) das System introjizierter Verhaltensnormen, sondern er beinhaltet - als eine Variante des nAchievement - : " the subject's highest hope, the dramatization of himself as a man of destiny". (a.a.O.,S.166)

Auf der Grundlage einer solchen Auswertung gilt es sodann, die 'Ebenen' festzustellen, welchen dieser sprachlich manifeste Gehalt der Testantworten zuzuordnen ist. Handelt es sich um einen latent need, um einen manifest need oder um eine Verhaltensnorm? Die Definitionen des Murray'schen need-Systems bieten auch für ein Kriteriumsverfahren, das für diese Aufgabe im folgenden zu konstruieren ist, eine brauchbare begriffliche Grundlage [1]. Dieses Verfahren wird im nächsten Abschnitt besprochen.

Kriterien für die Auswahl der in unserem Experiment verwendeten Variablen aus einem Gesamt von 28 need-Kategorien bei Murray wurden bereits im vorangehenden Abschnitt (s.S.126f) genannt. Sie sind im folgenden Abschnitt noch zu ergänzen.

---

1) Wir nehmen diese sehr umfangreichen Definitionen nicht in unseren Text mit auf. In einer abgekürzten Form können sie im ANHANG nachgeschlagen werden. Sie sind dort den items, die für jede Variable konstruiert wurden (Q-sort-Verfahren), vorangestellt. Für die vollständigen Definitionen sei auf Murray (1962) selbst verwiesen.

c) Die Q_s-sort-Methode als Kriteriumsverfahren

Es gilt nun, die hier konzipierte 'theoretische Realität' der Persönlichkeits-E b e n e n und ihrer gegenseitigen Relationen ('Ebenen-Gefüge') in eine 'experimentelle Realität' zu überführen [1]. Es sind also die Operationen zu bestimmen, welche die verschiedenen 'Erscheinungsmodi' (Ebenen) der oben zusammengestellten Variablen experimentell realisieren [2] und den Ausprägungsgrad einer 'Eigenschaft' (Variable) auf den verschiedenen Ebenen meßbar machen.

Lassen sich - auf dem Wege der Korrelation - signifikante Relationen nachweisen zwischen der durch diese Operationen definierten 'experimentellen Realität' und den T e s t - daten [3], so kann gefolgert werden : der Test repräsentiert die hier (probeweise) eingeführte 'theoretische Realität' [4], - allerdings nur in eben dem Maße, wie die experimentellen Operationen selbst diese 'theoretische Realität' repräsentieren. Denn die Gültigkeit eines Tests ist abhängig von der Gültigkeit des verwendeten Kriteriumsverfahrens.

Insbesondere für die 'Ebene der Motive' ist es schwierig, ein experimentelles Korrelat zu bestimmen. Zwar gibt es eine Reihe von Verfahren, welche ein bestimmtes Motiv 'messen' (vgl. u.a. die Arbeiten zur achievement-Messung, oder die Taylor-anxiety-scale). Doch ist - was nicht immer präzise gesehen wird - die ihnen zugeordnete 'theoretische Realität' auf die Bestimmungen der besonderen Verfahrensweise begrenzt, mit der Folge, daß verschiedene Verfahren sich nicht

---

1) Die weitere Diskussion beruht wiederum auf dem Ansatz von Holzkamp (1964), vgl. S. 99ff.
2) vgl. die Abgrenzung des Begriffs der 'experimentellen Realisation' von dem der 'operationalen Definition', s.S.112f.
3) Das Auswertungssystem, welches der Herleitung von Test-Daten vorausgeht, wird in Kapitel VII eingeführt.
4) vgl. die Auffassung eines Tests als 'experimentelle Realität' und seiner Validität als den Grad der 'Repräsentanz' zwischen Testdaten und theoretischer Realität, s.S.100.

miteinander vergleichen lassen. Ein weiteres Problem besteht darin, daß einerseits 'objektive' Meßverfahren keinen Aufschluß geben über die 'erlebten' Motive (welche wahrscheinlich höchst relevant sind für die Beurteilung des Verhaltens) und daß andererseits subjektive Äußerungen über die eigenen Motive nicht kongruent sein müssen mit jenen Antriebstendenzen, welche das Verhalten bestimmen.

Doch ist daraus nicht mit McClelland die Konsequenz zu ziehen, daß allein projektive Verfahren (Fantasie) ein angemessenes Instrument der Motivmessung darstellen, während die Selbstbeschreibung es nicht leiste "to reflect differences in motivation and only motivation" (1958, S.25). Vielmehr ist klarzustellen, daß es sich um grundsätzlich verschiedene Gegenstandskonstituierungen handelt, so daß man sich - je nach der theoretischen Fragestellung - für die eine oder andere zu entscheiden hat.

Fällt aber die Entscheidung auf den 'Erlebensaspekt' [1], so ist wiederum zu klären, ob auf dem Wege direkter Befragung tatsächlich die 'erlebten' Motive wiedergegeben oder vielmehr durch bewußtseinmäßige Zensur oder auch unkontrollierte Urteilseffekte entstellt werden. [2] Bei jenen Experimenten, welche mit direkter Motiv-Befragung arbeiten [3], werden bestimmte Bedingungen an solch direktes Vorgehen geknüpft, welche eine gewisse 'Gewähr' bieten sollen, daß derartige Entstellungen nicht zu erwarten sind. So soll bei den Vpn

---

1) Diese Entscheidung wird bei der Besprechung der Methode selbst eingehend begründet.
2) vgl. hierzu die sehr wichtige Unterscheidung zwischen 'Meinungen' (also hier 'Meinungen' über sich selbst) und 'Urteilen' bei Holzkamp (1964) und die Forderung, daß die Behauptung, die experimentell (z.B. anhand eines self-report) erhobenen 'Urteile' würden die theoretisch intendierten 'Meinungen' repräsentieren, erst einer Realisationskontrolle unterzogen werden müsse.
3) für Literaturangaben siehe Fußnote 2), S.136.

ein hinreichendes Maß an Intelligenz und Selbstreflektion gewährleistet sein : es wird durchgehend mit Studenten gearbeitet. Situative Determinanten sollen eine bewußte 'Zensur' so weit wie möglich unwirksam machen : durch Versuchs-'Atmosphäre' und Instruktion sollen die Vpn dazu geführt werden, 'offen', aufgabenorientiert und (etwa im Sinne eines wissenschaftlichen Interesses an der Fragestellung) so instruktionsgemäß wie nur möglich zu reagieren [1].

Die Schlußfolgerungen dieser Untersuchungen mit d i r e k t e r Motivbefragung sind jedoch nicht einheitlich [2]. Wir orientieren uns an den Arbeiten von Epstein (siehe Fußnote)), nach welchen das self-report - unter den Bedingungen einer entsprechenden Motivation zu 'offener' Selbstdarstellung - einen methodisch zulässigen Zugang zu den 'erlebten Motiven' darstellt und zwar dies auch für konfliktbesetzte Inhalte.

---

1) Bei Sherwood (1966) geht z.B. dem self-report eine eingehende Diskussion der betreffenden Variablen innerhalb eines psychologischen Kurses voraus. Davids und Pildner (1958) haben nachgewiesen, daß eine stress-situation eine angemessene Selbstdarstellung verhindert.

2) Davids et.al. (1955) entnehmen ihrer Untersuchung an Studenten, daß bei jenen Bedürfnissen, deren Realisation sozial zugelassen ist und welche entsprechend in einer Befragung verbalisiert werden können, die Selbstbeurteilung den 'projektiven' Zugang voll ersetzen könne. Auch Sherwood (1966) behält diese Möglichkeit ausschließlich den nicht-bedrohlichen Variablen vor, während Leimann und Epstein (1961) und Salz und Epstein (1963) mit Erfolg ein self-report über Konflikt und Schuldgefühle hinsichtlich sexuality und hostility erhoben haben. Eine wichtige Differenzierung scheint allerdings mit der (bereits erwähnten) Dimension 'revealing-conceiling' gegeben zu sein. Zimbardo (1964) z.B. weist nach, daß experimentell induzierte Furcht bei 'revealers' am besten durch ein self-report, bei 'conceilers' aber am besten durch projektives Material vorhergesagt wird. Diese Dimension hat hier nicht berücksichtigt werden können. Sie entspricht in etwa der Frage, ob und wieweit die auf dem Wege des self-report eingeholten 'Urteile' der Vpn tatsächlich ihr Selbstbild repräsentieren, welche hier nicht einer Realisationskontrolle unterzogen wurde (vgl. Fußnote 2) der Vorseite).

Es ist daher nun ein self-report zu konstruieren, das insbesondere dem Konzept der Ebenen Rechnung trägt und zugleich nach den hier interessierenden Variablen ausgerichtet ist. Diesen Erfordernissen kommt die Q-sort-Technik am nächsten.

Das Q-sort-Verfahren soll hier - im Rahmen einer allgemeinen Charakterisierung des experimentellen Vorgehens - nur so weit beschrieben werden, daß seine Rolle innerhalb des Experiments deutlich wird. Es wird also zunächst von den methodologischen Problemen dieses Verfahrens abgesehen; diese seien erst in dem sich anschließenden EXKURS (s.S.150) ausführlich diskutiert. Auch wird die Q-sort-Technik nur in der Form erörtert werden, welche auf Selbstbeurteilung (self-report) beruht. In Anlehnung an Cattell (1946) wird diese besondere Form im folgenden als Q_s-sort bezeichnet [1].

Das Q-sort-Verfahren ist zugleich mit allen self-report-Techniken dadurch charakterisiert, daß die 'Antworten' bereits vorgegeben sind und als solche dann irgendeine Art von Markierung oder quantitativer Einstufung durch den Probanden erfahren. Das Besondere der Q-sort-Technik aber liegt zu einem entscheidenden Teil gerade in der Weise des Einschätzens der items : es wird hier nicht jedem einzelnen item getrennt ein Punktwert seiner quantitativen Ausprägung zugeordnet; es wird auch nicht aus je verschiedenen item-Gruppen das relevanteste item ausgesucht (adjective-check-list, preference-test); sondern die items werden - nicht in Listen, sondern je einzeln auf Karten präsentiert - von den Vpn nach dem Grade ihres 'A u f - m i c h - z u - t r e f f e n s' beurteilt. Dieser graduellen Abstufung

1) Das Signum 's' verweist auf den 'solipsistischen' Charakter der Daten. Cattell führt diesen Begriff ein, um die Tatsache zu kritisieren, daß es sich bei diesem Vorgehen um 'Introspektion' handelt (dazu ausführlicher im Exkurs); vgl. den weniger präzisen Ausdruck 'S-sort' bei Mowrer (1953).

nach werden sie 's o r t i e r t' [1] und zumeist entlang einer mehrstufigen Skala in eine angenäherte Normalverteilung ausgelegt. [2] Der Vorzug einer solchen Verfahrensweise wird deutlich durch den nächsten Gesichtspunkt.

Eine weitere Charakteristik nämlich besteht in der durch das 'Q' bezeichneten Korrelationstechnik. [3] Im Gegensatz zur 'R'-Technik werden hier verschiedene Äußerungsformen e i n e s Subjekts miteinander korreliert, eine Technik, die sich für die Zwecke dieser Arbeit, nämlich für einen Zugang zum Ebenen-Gefüge e i n e r Person, hervorragend eignet.

Ein solches Verfahren erlaubt es nun, unter allen nur denkbaren Äußerungen einer Person, insbesondere die Ich-Ideal-Vorstellungen (ideal-self) und das Selbstbild (real-self), nämlich die Beurteilung des eigenen V e r h a l t e n s , zu erheben und den Grad ihrer gegenseitigen Beziehung quantitativ zu bestimmen.

Es werden hier anhand derselben items sowohl das eigene Verhalten (self-sort), wie das Ich-Ideal (ideal-sort) eingeschätzt, indem die items jeweils dem Grade ihres Zutreffens nach am Maßstab einer vorgegebenen Skala 'sortiert' werden : es läßt sich sodann für jedes item die Differenz der Skalenpunkte im real- und ideal_ sort, sowie die Summe aller Differenzen bestimmen, oder - eine forced-distribution vorausgesetzt - der Korrelationskoeffizient zwischen beiden Selbstbeschreibungen auf der Grundlage gleichzeitig aller items berechnen. Damit ist ein quantitatives Maß gegeben für den

---

1) daher die Bezeichnung Q - 's o r t' .
2) Es wird hier nur die Form der 'forced-distribution' (in der also eine bestimmte Verteilungsform vorgeschrieben wird) dargestellt, da sie den praktischen Erfordernissen dieser Arbeit am meisten entspricht; ihre methodologischen Probleme werden im Exkurs diskutiert.
3) Die Bezeichnung 'Q'-Korrelation wurde von Stephenson und Thomson 1935 (unabhängig voneinander) eingeführt. Sie bezeichnete damals auch die Korrelation von Personen mit Personen zur Aufstellung von Typen, wie auch zur Messung der Reliabilität von 'Urteilern'.

Grad der Übereinstimmung (Differenz) zwischen beiden Formen der Selbsteinschätzung, der Einschätzung des eigenen Verhaltens (real-self) und des eigenen Verhaltens-Anspruchs (ideal-self). [1)]

Sieht man einmal vom introspektiven Charakter der Daten ab, so ist mit der Q_s-sort-Technik eine Methode gegeben, mit welcher sich verschiedene 'Ebenen' (real-self, ideal-self) sowohl getrennt erheben, wie auch quantitativ miteinander vergleichen lassen. Das Verhältnis zwischen unserem Begriff der 'Persönlichkeitsebene' und einer derartigen Operation wird im folgenden zu klären sein. Doch sei zunächst - in Analogie zu den zwei beschriebenen Formen der Selbsteinschätzung - noch ein weiterer Begriff eingeführt.

Wir haben oben (s.S.134ff) diskutiert, daß es methodisch zulässig sei, die 'erlebten Motive' auf dem Wege der Selbsteinschätzung zu erheben. Es bietet sich daher an, auch den Begriff der 'Ebene der Motive' in Form eines Q_s-sort's (also in Form einer Selbstbeschreibung) zu 'realisieren'. Das bedeutet : die Vp hat mit denselben items nicht nur ihr Verhalten (real-self) und ihre Verhaltensnormen (ideal-self) einzuschätzen, sondern auch ihre 'Bedürfnisse'. Sie hat zu charakterisieren, wie sie sich a n t r i e b s m ä ß i g am liebsten verhalten würde : für den fiktiven Fall, daß solches Verhalten ungeachtet der sozialen Umwelt möglich, daß die Anpassung an die Bedingungen der Realität nicht erforderlich sei. Dieses Bezugssystem für die Selbstbeurteilung nennen wir fortan 'Wunsch-Ich'. [2)]

Den 'theoretischen' Begriffen Verhalten, Antrieb und Verhaltensnorm entsprechen somit die 'experimentellen' Begriffe real-self, Wunsch-Ich und ideal-self. [3)] Dem theoretischen

---

1) Auf die Fülle der Arbeiten, in welchen die Q_s-sort-Technik in der beschriebenen Weise verwandt wird, wird im EXKURS im einzelnen verwiesen (s.S.150ff und Fußnote 1),S.177).

2) Wir behalten die englischen Bezeichnungen real-self und ideal-self bei, da sie sich in der Literatur als Fachtermini eingebürgert haben. Der Begriff 'Wunsch-Ich' wird, da er hier neu eingeführt wurde, durch die deutsche Fassung von diesen abgehoben.

3) Es sei angemerkt, daß diese Begriffe in der angelsächsischen Literatur nicht durchgängig als rein operationale Begriffe verwendet werden.

Begriff der 'Persönlichkeitsebene' ist der operationale Begriff der 'Urteilsebene' oder der der 'Bezugsebene' zugeordnet. Der theoretische Begriff des 'Ebenen-Gefüges' wird realisiert durch die quantitativ faßbaren Relationen zwischen diesen verschiedenen 'Urteilsebenen'.

Es werden also Q_s-sort-items zu konstruieren sein, für welche sich der Grad ihres 'Zutreffens-auf-mich' sowohl für das eigene Verhalten, wie für das Ich-Ideal, wie für das Wunsch-Ich bestimmen, d.h. auf einer mehrstufigen Skala quantitativ markieren läßt. Zugleich müssen die items inhaltlich ausgerichtet sein an den bereits für diese Untersuchung zusammengestellten Variablen. Die Aufgabe der Vpn besteht demnach darin, einen bestimmten (später einzuführenden) Satz von Q_s-sort-items dreimal zu 'sortieren', nämlich die items in drei verschiedenen Urteils-Bezugssystemen dem Grade ihres Zutreffens oder ihrer 'intra-individuellen Signifikanz' (Stephenson 1953) nach einzuschätzen.

Das zunächst allgemein aufgrund persönlichkeitstheoretischer Erwägungen (und experimenteller Ergebnisse hinsichtlich projektiver Tests) entwickelte Ebenen-Konzept kann im weiteren nur in den Grenzen diskutiert werden, die durch diese experimentellen Bedingungen gesteckt sind. Orientiert man sich an dem präzisen Kriterium, das durch den Begriff der 'Repräsentanz' gegeben ist, so lassen sich diese Grenzen des Aussagewertes der hier vorgeschlagenen Operation (Q_s-sort) genau bestimmen.

Entscheidend für das Verhältnis von theoretischem Konzept und experimenteller Vorgehensweise ist, daß die 'experimentelle Realität' nur durch 'Lebensraumbestimmungen' gekennzeichnet ist. Das bedeutet, daß auch jene 'theoretische Realität', innerhalb welcher das Experiment zu interpretieren ist, nur durch entsprechende 'Lebensraumbestimmungen' definiert sein kann. Es kann also im weiteren nicht von d e n Motiven, von d e m Verhalten gesprochen werden, sondern nur von dem subjektiv-vorstellungsmäßigen Abbild dieser 'Ebenen' oder von Selbstwahrnehmungen bzw. Selbstbeurteilungen.

Genau genommen aber ist auch eine solche Interpretation noch in Zweifel zu ziehen; denn es ist - darauf war schon hinge-

wiesen worden - nicht geklärt, ob das hier verwendete Verfahren garantiert, daß diese Selbstwahrnehmungen tatsächlich adäquat wiedergegeben und nicht durch selegierende oder modifizierende Urteilseffekte entstellt werden. Wir können hier nur voraussetzen, daß sich derartige störende Effekte hinreichend dadurch ausschalten lassen, daß durch die Versuchs-Situation eine entsprechende Motivation zu instruktionsgemäßer Reaktion induziert wird [1]. Diese Annahme, daß man von derartigen Störeffekten absehen könne, bezieht sich nicht - das sei noch einmal eigens betont - auf die Frage, ob die 'Urteile' den o b j e k t i v e n Tatsachen entsprechen, sondern allein darauf, ob die subjektiven Meinungen adäquat wiedergegeben werden; es geht nicht um das 'objektive' Motiv, sondern um das 'erlebte' Motiv, nicht um das 'tatsächliche' Verhalten, sondern das Selbstbild der Vp.

Es gilt nun aufzuzeigen, daß diese Art der Realisation der theoretischen Begriffe nicht eine Einschränkung für das theoretische Konzept der 'Persönlichkeitsebene' bedeutet, sondern daß der 'Erlebensaspekt' einen notwendigen Bestandteil der Definition jener 'theoretischen Realität' darstellt, um welche es speziell in dieser Untersuchung nur gehen kann.

Wenn es, unserem Ansatz nach, für die Art der Test-Reaktion ausschlaggebend ist, ob ein Bedürfnis angst- oder schuldbesetzt, ob es befriedigt oder den eigenen Normen widersprechend ist, so sind hier Lebensraumtatbestände gemeint. Jenes Schuldgefühl, welches den Ausdruck aggressiver Tendenzen hindert, ist ein 'erlebtes' (sogar 'erlittenes') Schuldgefühl. Die Differenz zwischen Verhaltensanspruch und seiner Realisation ist gerade dann (u.U.) ausschlaggebend für die Test-Reaktion, wenn diese Differenz 'erlebt' wird, wenn die Vp sich als jemand erfährt, der den eigenen Normen nicht entspricht.

---

1) Vollständig kontrollierte Versuchsbedingungen lassen sich allerdings im Rahmen einer solchen Aufgabenstellung wohl grundsätzlich nicht herstellen, da die sozialpsychologische Situation des Experimentes immer noch inter-individuell verschieden e r l e b t werden kann, was sich eben modifizierend auf die 'Urteile' auswirken würde.

Das methodische Problem besteht also nicht darin, daß der 'Lebensraumaspekt' z u g e l a s s e n wird; sondern der methodisch kritische Punkt liegt darin, daß für derartige Selbstaussagen zunächst festzustellen ist, ob sie g ü l t i g sind, ob sie tatsächlich ein adäquates Abbild des 'subjektiven Erlebens' vermitteln. Auf diese Frage nach der Validität des verwendeten Q_s-sort-Verfahrens kommen wir weiter unten zurück.

Zunächst ist auf Einschränkungen zu verweisen, die mit dieser Form der experimentellen Realisation des Ebenen-Begriffs verknüpft sind. Wir haben nicht die Möglichkeit, 'unbewußte' Vorgänge in unsere Betrachtung mit einzubeziehen oder die - theoretisch weniger belastete - Frage nach der Rolle der 'Selbsteinsicht' zu stellen. Denn wir haben kein unabhängiges Kriterium, nach welchem sich die Differenz zwischen 'erlebtem Verhalten' und 'tatsächlichem Verhalten' oder zwischen 'erlebtem Motiv' und 'tatsächlichem Motiv' bestimmen ließe. Wir müssen damit eine vielleicht wichtige Frage aus unserer Untersuchung ausklammern : reagieren Vpn, die ihr Verhalten falsch einschätzen und daher eine bei ihnen 'objektiv' vorliegende Differenz z.B. zwischen Verhaltensnorm und Verhalten nicht erfahren, anders auf Test-stimuli als Vpn, welche diese Differenz 'erleben'? Oder reagieren beispielsweise Vpn, welche ihr starkes Bedürfnis, bewundert zu werden, fälschlich als 'befriedigt' erfahren (weil sie die Umwelt falsch wahrnehmen, z.B. im Sinne einer autistischen Projektion, vgl.S. 17) anders auf entsprechende stimuli als Vpn, welche bei gleich starkem Bedürfnis dieses als unbefriedigt erfahren, indem sie die Umwelt (resp. die Reaktionen der Beziehungspersonen) richtig einschätzen? [1] Wir können nicht entscheiden, ob auch solche Differenzen, die nicht 'erlebt' werden, eine Rolle für das Testverhalten spielen.

---

1) vgl. S. 16 , Fußnote 1) und S. 57

Das bedeutet : die Trennkriterien, die wir vermittels des Ebenen-Konzeptes gewinnen wollen, jene Kriterien also, die Gruppen von Vpn mit systematisch unterschiedlicher Testreaktion differenzieren sollen, sind allein durch 'Lebensraumbestimmungen' definiert. Die Form der Operationalisierung, die wir für das theoretische Konzept der 'Ebenen' wählten, impliziert : allein, wie eine Vp sich e r l e b t , wird zum Kriterium für die Art der zu erwartenden Testreaktion. Indes ist es kaum wahrscheinlich, daß im Bereich des 'Erlebens' allein die für die Testreaktion relevanten inter-individuellen Differenzen zu suchen sind.

Wir kehren nun zur Frage nach der Validität des verwendeten Q_s-sort-Verfahrens zurück. Würde man es als garantiert ansehen können, daß die Antworten in diesem self-report ein zutreffendes Abbild der 'Meinungen' der Vp über sich selbst liefern, daß also die Reaktion auf ein bestimmtes item tatsächlich die Selbstwahrnehmung i.b. auf das in dem item angesprochene Verhalten zutreffend wiedergibt, so ist damit eine entscheidende Frage noch nicht geklärt : die Frage nämlich, wofür diese Art der Selbstwahrnehmung ihrerseits 'repräsentativ' ist. Die im self-report (Q_s-sort) verwendeten items stellen ja in jedem Falle nur einen (äußerst schmalen) Ausschnitt aller nur denkbaren Aussagen (statements) über das Verhalten dar. Beantwortet z.B. eine Vp alle items, die Aggressivität betreffen, positiv, so bleibt zu fragen : ist diese Vp durchgehend stark aggressiv oder ist sie aggressiv nur in jenen sozialen Situationen, die in den betreffenden items beschrieben werden. Oder schätzt eine Versuchsperson das item 'Ich bin sehr unglücklich, wenn ich nicht unter Menschen bin' als sehr zutreffend ein, so läßt sich nicht ohne weiteres folgern, die Vp sei durchgängig durch ein hohes 'nAffiliation' charakterisiert. Zwar wird nAff u.a. durch eben dieses item definiert : doch sind alle items, welche in einem bestimmten Verfahren zur Beschreibung einer einzigen Variablen konstruiert werden, ihrerseits nur ein Bruchteil aller denkbaren Beschreibungen des betreffenden Verhaltens.

Damit stehen wir vor dem Problem der Repräsentativität der verwendeten item-Stichprobe. Dieses Problem und sein Zusammenhang mit der Gültigkeitsbestimmung eines Q_s-sort-Verfahrens wird im EXKURS ausführlich erörtert. Hier sei nur vermerkt, daß in keiner der bisher veröffentlichten Arbeiten ein befriedigender Nachweis der Validität des jeweils verwendeten Q_s-sort-Verfahrens erbracht wurde (vgl.S. 177).

Wir kommen im EXKURS - für eine ausführliche Diskussion sei auf diesen selbst verwiesen - zu folgender Lösung. Die items für ein Q_s-sort sind nicht aus einem gedachten 'item-Universum' als eine 'repräsentative Stichprobe' zu s e l e g i e r e n; sondern die items sind - innerhalb des festgelegten Rahmens einer bestimmten Theorie - zu k o n s t r u i e r e n . Sie sind so zu konstruieren, daß - nunmehr im Holzkamp'schen Sinne der 'Repräsentanz' - die experimentelle Operation die vorgeordneten theoretischen Begriffe 'realisiert' und für diese 'repräsentativ' ist. Wir haben allerdings diese Operation (Q_s-sort) nicht einer umfangreichen Realisationskontrolle unterzogen : die empirische Validität des Verfahrens ist nicht bestimmt worden. Stattdessen sind einzelne theoretische Annahmen über Interrelationen der Daten (Einschätzungen) des Q_s-sort-Verfahrens hergeleitet worden, deren Bestätigung als ein Beitrag zur construct-Validität dieses Verfahrens gewertet werden kann. (s.S. 184ff) Diese theoretischen Annahmen ergeben sich aus den Definitionen der verschiedenen 'Ebenen' : ihre Bestätigung beinhaltet, daß entscheidende Realitätsbestimmungen des theoretischen Konzeptes als experimentell realisiert angesehen werden können. Wir verweisen hier auf den experimentellen Teil (s.S.206ff), wo aufgezeigt wird, daß das von uns konstruierte Q_s-sort-Verfahren derartigen theoretischen Kriterien entspricht.

Die V a r i a b l e n , welche durch das Q_s-sort-Verfahren repräsentiert werden sollen, wurden bei der Besprechung der WP (s.S.126f) hergeleitet. Für dieselben Variablen, für welche angenommen wird, daß sie in den Reaktionen auf die WP auftreten, soll ja das Selbstbild der Vp erhoben werden. Der

Ausdruck etwa von nAgg in der WP soll verglichen werden mit den 'Erscheinungsmodi' dieser Variablen auf den verschiedenen Ebenen, soll bezogen werden auf das 'Ebenen-Gefüge' hinsichtlich dieser Variablen [1]. Eben dieses Ebenen-Gefüge wird durch das Q_s-sort repräsentiert. [2]

---

1) Es sei daran erinnert : das Ebenen-Modell beschreibt die Repräsentanz e i n e r bestimmten Variablen auf den verschiedenen Ebenen (vgl.S.116f).

2) Hier sei angemerkt, daß der hier unternommene Versuch, die Q-sort-Technik auf Fragen der projektiven Testdiagnostik anzuwenden, nicht der erste ist. Allerdings gibt es offenbar bisher nur eine Arbeit, welche - unserem Vorgehen ähnlich - den Motivausdruck einer bestimmten V a r i a b l e n (hier nAchievement) im projektiven Test vergleicht mit der Selbsteinschätzung dieser Variablen und insbesondere mit dem real-self/ideal-self discrepancy-score (Martire,1958). Zumeist wird der Vergleich zwischen projektivem Test und Selbsteinschätzung anhand eines weiteren Q-sort's vorgenommen, welches ein 'Urteiler' auf der Grundlage des T e s t -Materials über die betreffende Vp zusammengestellt hat : es wird also die Testinterpretation in Form eines Q-sort's formuliert und dieses wird dann mit dem self-report der Vp (welchem dieselben items zugrundeliegen) verglichen (vgl. Friedman 1955, Filmer-Bennet und Klopfer 1962). Ähnlich werden auch Diagnosen von Psychiatern, die sich nicht auf Tests, sondern auf ein psychiatrisches Interview stützen, in Form eines Urteiler-Q-sort's ausgedrückt und auf das self-sort des Patienten bezogen (Kogan et.al. 1957). Solches Vorgehen gründet jedoch auf dem Prinzip der Einheit von Test und Interpret bzw. im letzteren Falle auf der von Diagnose und Psychiater und erschließt somit gerade nicht einen 'objektiveren' Zugang zum Validitätsproblem. Dasselbe gilt für die Arbeit von Silverman (1959) : er vergleicht die Urteiler-Q-sort's (die hier auf Test- und Interviewdaten basieren) von Psychologen und Psychiatern verschiedener Provenienz, welche zugleich mehr oder weniger 'klinische Erfahrung' haben. Auch diese Anwendung der Q-Technik bei Silverman scheint uns - im Gegensatz zu Hörmann (1964a) - keinen Fortschritt gegenüber den verschiedenen Zuordnungsmethoden darzustellen, bei welchen auf einem nur weniger streng quantitativen Wege gleichfalls verschiedene 'Gutachten' über eine Vp (zumeist blind) einander zugeordnet werden (matched). Es scheint uns hier vielmehr ein anderes Argument von Hörmann angebracht zu sein, nämlich, daß es genauso unmöglich sei, an der Validierung eines Tests auf der Basis 'Test-Interpret' den Wert des Tests abzulesen, wie an ihr den Wert des Diagnostikers zu bestimmen (1964a, S.101).

Diese Auswahl der Variablen ist jedoch für die Konstruktion eines Q_s-sort-Verfahrens aus methodischen Gründen zu erweitern. Die Aufgabe der Vp besteht - das sei noch einmal vergegenwärtigt - darin, alle vorgelegten items entlang einer mehrstufigen Skala, welche ein Kontinuum zwischen dem Grad höchsten Zutreffens bis zur scharfen Ablehnung beschreibt [1], zu sortieren. Die Vp hat sich dabei an die vorgeschriebene Verteilungsform einer angenäherten Normalverteilung zu halten (forced method) [2]. Wenn aber die Vp nicht frei ist, z.B. das positive oder das negative Ende der Skala zu bevorzugen, sondern gezwungen ist, alle Skalenpunkte in vorgeschriebenem Maße zu benutzen (forced method), so muß folgendes gewährleistet sein : die items müssen vielfältig genug zusammengestellt sein, damit jede Vp sowohl positive, wie negative Wahlen treffen kann [3]. Würden in einem Q_s-sort etwa nur sozial sehr positiv bewertete 'Eigenschaften' beschrieben, wäre es für die Vp kaum möglich, z.B. für das ideal-self die ganze Skala zu verwenden : bei einer 'freien Verteilung' würde nur die positive Seite der Skala benutzt.

Wir haben daher die in Orientierung an den erwarteten Aussagemöglichkeiten der WP zusammengestellten Variablen durch 'komplementäre' Variablen ergänzt : nAggression wurde ergänzt durch nAbasement, nDominance durch nDeference, nNurturance durch nSuccorance, nAffiliation durch nRejection. Den Variablen nAutonomy, nExhibition und Harmavoidance wurden keine weiteren hinzugefügt.

---

1) Die negativen Extrempunkte der Skala beinhalten, daß man sich durch das betreffende item überhaupt nicht charakterisiert, daß 'eher das Gegenteil zutrifft' (vgl. die Instruktion zum Q_s-sort-Verfahren, s. ANHANG, S. 334).
2) Im EXKURS wird ausführlicher diskutiert, warum mit der forced method gearbeitet wird (s.S.157ff) und warum als Verteilungsform eine angenäherte Normalverteilung gewählt wird (s.S.166f).
3) siehe Fußnote 1).

Mit einer solchen Auswahl von Variablen ist nicht mehr bezweckt, als daß die items, die eine Vp - in vorgeschriebener Abstufung - einschätzen soll, inhaltlich weit genug streuen. Es ist damit nicht die Annahme verknüpft, daß z.B. Vpn, die nDom hoch einschätzen, nDef gering einschätzen müßten, daß also logisch gegensätzliche 'Eigenschaften' sich auch psychologisch ausschließen müßten : durch das item-Angebot kann nicht garantiert werden, daß die Einschätzungen der items sich 'normal' verteilen. [1] Es sollen nur gewisse Vorbedingungen dafür geschaffen werden, daß die vorgeschriebene Verteilungsform nicht in erheblichem Maße von jener Verteilungsform abweicht, welche bei freiem Sortieren entstände (dazu ausführlicher im EXKURS, s.S. 166).

Die D e f i n i t i o n e n dieser Variablen übernehmen wir - wie ausgeführt (s.S.130ff) - von Murray. Die Aufgabe der Konstruktion eines Q_s-sort-Verfahrens besteht also darin, diesen Definitionen entsprechend einzelne items zu formulieren. Dabei sollte jede Variable durch möglichst viele verschiedene items vertreten sein, damit ein möglichst weiter Ausschnitt verschiedener Verhaltensmanifestationen (z.B. für nAgg) durch das Verfahren gedeckt ist. Nur eine hinreichende Vielfalt der in den items beschriebenen Situationen macht es möglich, daß jede einzelne Vp ihr eigenes Verhalten nach Situationen differenzieren kann und daß so auch inter-individuelle Differenzen hinsichtlich einer Variablen deutlich genug in Erscheinung treten.

Die items werden in Form von Persönlichkeits- s t a t e m e n t s formuliert. Für die Konstruktion derartiger Persönlichkeits-statements haben wir einige Kriterien zusammengestellt, die sich auf Inhalt und Form der Formulierung beziehen. Sie sind, zusammen mit den items selbst, im ANHANG (s.S. 323) aufgeführt.

---

1) Auf die Frage der Normalverteiltheit von Eigenschaften im Individuum selbst gehen wir im EXKURS ein (s.S. 168f).

Jeder der 11 Variablen (s.S.146) sind 8 items zugeordnet. Diese item-Anzahl ist als eine Art Kompromiß zu verstehen : einerseits sollten möglichst verschiedenartige Situationen beschrieben werden (s.o.); zum anderen aber ergab sich aus den ersten Vorversuchen, daß eine zu hohe Gesamtzahl von items (im Vorversuch waren es 120 items) durch die Vpn nicht 'sortiert' werden konnte. Die Aufgabe, derart viele items in der vorgeschriebenen Form einer angenäherten Normalverteilung zu sortieren, d.h. jedem item in Abwägung der übrigen items einen bestimmten Ort auf der vorgelegten Skala zuzuweisen, führte zu schnell zu Ermüdung bzw. Nachlassen der Konzentration. Wir haben daher die Gesamtzahl auf 88 items reduziert.

An der item-Anzahl war die Form der S k a l a , wie auch die (vorgeschriebene) Antwortenzahl pro Skalenpunkt auszurichten. Die günstigste Verteilung (in Form einer abgeflachten Normalverteilung, vgl. EXKURS, S. 166) ergibt sich bei einer 9-Punkte-Skala. Diese Skala wurde so gewählt, daß positive und negative Wahlen anschaulich voneinander abgehoben sind, nämlich in der Form : +4,+3,+2,+1,0,-1, -2, -3, -4. Der Skalenpunkt +4 bezeichnet den Grad höchsten Zutreffens, -4 den Grad stärkster Ablehnung, der Skalenpunkt 0 beinhaltet Indifferenz gegenüber der in einem item gegebenen Persönlichkeitsbeschreibung. Die Anzahl der items, die jedem Skalenpunkt zuzuordnen war, läßt sich aus folgender Übersicht ersehen :

+4 (5), +3 (8), +2 (10), +1 (13), 0 (16), -1 (13), -2 (10), -3 (8), -4 (5).

Es sei noch einmal zusammengefaßt : das Q_s-sort enthält für 11 Variablen je 8 items. Die Vp hat also für 88 items den Grad des 'Zutreffens auf mich' einzuschätzen, indem sie die vorgelegten Karten (jedes item ist auf eine Karte gedruckt, vgl. S. 191) nach obiger Vorschrift 'sortiert'. Derselbe Satz von items ist dreimal zu sortieren : einmal ist anzugeben, ob die vorgelegten Beschreibungen das eigene Verhalten zutreffend charakterisieren (real-self), zum anderen ist anzugeben, ob das beschriebene Verhalten den eigenen Normen entspricht (ideal-self) und schließlich

ist einzuschätzen, inwieweit man sich in der angegebenen Weise verhalten würde, stünden dem keinerlei Hindernisse entgegen (Wunsch-Ich).

Die Probleme, die sich aus dieser Methode ergeben, werden in dem sich anschließenden EXKURS diskutiert. Das Datenmaterial, das sich diesem Verfahren entnehmen läßt, wird im experimentellen Teil im einzelnen beschrieben (s.S.202ff).

VI. E x k u r s : Die Q-sort-Methode in ihren theoretischen und experimentellen Implikationen

1. Vorbemerkungen

'Q - t e c h n i q u e' ist insofern ein mehrdeutiger Begriff, als diesem verschiedene Techniken subsumiert werden und verschiedene Autoren dieser Methode einen sehr unterschiedlichen Bedeutungsradius zuschreiben. Das Zeichen 'Q' (im Gegensatz zu 'R') steht sowohl für die Korrelation von Personen, wie von verschiedenen items hinsichtlich einer Person; es steht für die Korrelation von solipsistischen scores, wie von ipsativen Daten [1], wie auch von Messungen objektiver Tests. Es steht für eine mögliche Korrelations- t e c h n i k unter anderen, wie für eine umfassende M e t h o d o l o g i e , einschließlich deren Voraussetzungen an psychologischer Theorie und wissenschaftstheoretischer Orientierung (so bei Stephenson 1953).

Der Begriff 'Q - s o r t' läßt sich dadurch näher spezifizieren, daß es sich hier in jedem Falle um U r t e i l e handelt, die in einer bestimmten festgelegten Form zum Ausdruck zu bringen sind: die Urteile selbst nämlich sind (auf Karten) vorgegeben und werden anhand einer ebenfalls vorgegebenen Skala nur ihrem relativen Gewicht für die jeweilige Fragestellung nach 'sortiert'. Der Gegenstand solchen Urteilens ist beliebig; es kann sich um Personen, wie um irgendwelche konkreten oder auch ideellen Objekte handeln. Dem Q-sort können somit jederart 'subjektive Einschätzungen' zugrundeliegen, sofern sie sich um einen gedachten Mittelpunkt gruppieren lassen. [2]

Der Begriff '$Q_{\underline{s}}$ - sort' kennzeichnet allein S e l b s t - einschätzungen, also introspektive Urteile. Da allein diese Form in dieser Arbeit interessiert, ist auch nur diese ein-

---

1) Es sei noch einmal erinnert, daß dieser Begriff 'solipsistisch' in Anlehnung an Cattell (1944,1946) verwendet wird: es handelt sich um ipsativ um einen intra-individuellen Mittelpunkt angeordnete introspektive Daten.

2) Es ist hier also nicht die Q-Korrelation von Verhaltens- oder objektiven Testdaten mit umfaßt.

gehender zu behandeln. Es trifft sich allerdings, daß dieser Form einer Anwendung der Q-technique die meisten und schwerwiegendsten Probleme anhaften. Doch auch sie seien nur so weit diskutiert, wie sie für das hier durchgeführte Experiment relevant sind. Sie sind durchgängig dadurch bestimmt, daß solipsistische Daten einer statistischen Manipulation unterzogen werden sollen.

Soweit die Argumentation durch Beispiele veranschaulicht wird, sind diese dem für die Zwecke dieser Arbeit konstruierten Q_s-sort entnommen. Denn es soll hier weniger ein 'Sammelreferat' in Orientierung an der in der Literatur geführten Diskussion gegeben, als vielmehr das experimentelle Vorgehen allein dieser Arbeit so weit wie möglich begründet werden.

2. Zum Problem solipsistischer Kontinua.

Bevor alle weiteren methodologischen Implikationen der Q_s-sort Methode betrachtet werden können, muß als Ausgangsfrage geklärt werden, ob denn die Aufgabe, welche dieses Verfahren an die Vp oder den Pb stellt, überhaupt ausführbar ist. Die schärfsten Kritiker dieser Methode nämlich verneinen diese Frage bzw. geben jedenfalls ihren theoretischen Gegenargumenten ein zusätzliches Gewicht durch die Behauptung, daß diese Aufgabe kaum realisierbar sei (z.B. Cattell 1952). Die Schwierigkeit dieser Aufgabe liegt darin, daß sehr differente Aspekte auf einem gemeinsamen Kontinuum angeordnet werden sollen. Sie wird besonders deutlich, wenn man die Q_s-sort-Skala z.B. mit einer Einstellungsskala vergleicht. Hier nämlich, in der Einstellungsskala, kommt jedem einzelnen item (jeder Frage) sozusagen eine eigene Dimension zu; die Einschätzung eines jeden items ist (im Prinzip) unabhängig von der Einschätzung jedes anderen items. Soll man z.B. die Einstellung zu einem Satz wie 'Kriege sind notwendig' quantitativ (als Grad der Übereinstimmung mit diesem) markieren, so wird ausschließlich im Bezugssystem möglicher Einstellungen zu der in diesem item angegebenen Meinung geurteilt, ein Bezugssystem also, das samt seiner Urteilsskala allein diesem bestimmten item zukommt. Oder soll anhand einer rating-Skala die Aggressivität einer anderen Person

beurteilt werden, so vollzieht sich auch dieses Urteil allein im Bezugssystem 'Aggressivität', d.h. im Rahmen der (gedachten) Variationsbreite möglicher Ausprägungsgrade dieser bestimmten Verhaltensdeterminanten : das zwar subjektive Urteil kann an einem (wenn auch wiederum subjektiv bestimmten) n o r m a t i v e n Maßstab ausgerichtet werden.

Im Q_s-sort-Verfahren entspricht dagegen nicht jedem item eine allein diesem zugehörige Dimension : es sollen vielmehr alle noch so heterogenen Einzelurteile (items) in eine einzige ihnen gemeinsame Dimension gebracht werden. Theoretisch läßt sich argumentieren, daß zwar derart heterogene Persönlichkeitsaspekte nicht als solche vergleichbar sind, daß sie jedoch im Bezugssystem einer einzigen Person sich vereinheitlichen lassen durch die jeweilige Bedeutung, durch das jeweilige Gewicht, das sie hier einnehmen. So spielt etwa 'Hilfsbereitschaft' im Verhalten einer bestimmten Vp eine größere Rolle als 'Neid' gegenüber anderen Menschen. Im Vergleich zueinander haben solche verschiedenen 'Eigenschaften' eine durchaus verschiedene 'subjektive Relevanz' : sie charakterisieren sie mehr oder weniger deutlich, individualisieren sie in einem je verschiedenen Maße. Das Individuum selbst hat - aufgrund seiner Erfahrung mit sich selbst und mehr oder weniger deutlich ausgeprägt - ein subjektives Bild über die verschiedenen Grade der Ausprägung solcher Verhaltensaspekte und kann sie auch gerade i n dieser Heterogenität nach dem Grade ihrer jeweiligen 'intra-individuellen Signifikanz' (Stephenson 1953) oder in ihrer 'likeness to me' (Broverman 1962) urteilsmäßig erfassen. Ob aber nun diese erlebte intra-individuelle Signifikanz verschiedener 'Eigenschaften' vom Subjekt auch q u a n t i t a t i v wiedergegeben, also 'geschätzt' werden kann, ist damit noch nicht entschieden.

Stephenson bejaht diese Frage, wenn er argumentiert, daß als eine wichtige Bedingung für die Durchführung von Korrelationen hier, auf der Dimension der 'intra-individuellen Signifikanz', das 'transitory postulate' [1] ebenso zutreffe, wie

1) Die Bedingung nämlich, daß wenn x>y und y>z, dann auch x>z. Ob diese Bedingung für die Durchführung von Korrelationen aber hinreichend ist, wird an späterer Stelle noch einmal zu diskutieren sein (s.S.171).

es bei jenen Skalen gilt, welche der R-Korrelation zugrundeliegen. Doch setzt auch dieses theoretische Postulat seinerseits die praktische Realisierbarkeit eines solchen solipsistischen Kontinuums voraus. Gerade dieses aber wird durch Cattell (1952) und auch durch Travers (1951) in Frage gestellt.

Es soll im folgenden eine beiden Autoren entsprechende Lösung vorgeschlagen werden : doch wird sie erst zureichend begründbar, wenn allererst der Umgang mit einer Q_s-sort-Skala (oder die Reaktion auf diese, gefaßt als Prozeß) psychologisch betrachtet wird.

Jede Selbstbeurteilung nämlich setzt, wie jede Persönlichkeitsbeurteilung, die Orientierung an irgend einer Norm voraus : die Stärke meiner Aggressivität erkenne ich nur durch den Vergleich mit 'anderen'; das relativ geringe Charakterisiertsein durch Dominanz erschließt sich ebenfalls nur dadurch, daß meine Art von dominantem Verhalten mich nicht in einer solchen 'outstanding' Form, wie es für Aggressivität der Fall ist, vom Durchschnitt meiner Bezugsgruppe unterscheidet. Erst in solchem Vergleich zu einer (oder je nach der 'sozialen Rolle' mehrerer) Bezugsgruppe (n) kann überhaupt ein 'Selbstbild' entstehen.

Dabei muß diese Orientierung am Maßstab einer Bezugsgruppe sich nicht bewußt oder explizit vollziehen. Vielmehr werden im Verlaufe der Sozialisation derartige Beurteilungsnormen erworben, so daß sie indirekt Selbst- und Fremdbeurteilung bestimmen. Allerdings sind diese Beurteilungsnormen nicht nur inter-individuell verschieden, sondern können auch intra-individuell variieren, so daß für die Zwecke eines Experiments auch aus diesem Grunde eine einheitliche Norm gesetzt werden muß. Es sei dies im folgenden an einem Beispiel veranschaulicht, welches zugleich deutlich machen soll, daß auf diesem Umwege über eine inter-individuelle Norm zumindest ein intra-individuelles Kontinuum, nämlich die Beurteilung heterogener Verhaltensaspekte auf einer gemeinsamen Dimension, tatsächlich realisierbar ist.

Soll ich mich z.B. hinsichtlich der subjektiven Relevanz von nAgg, nDom und nNur beurteilen, so sind,in vereinfachter Form,zwei Extrem-Fälle zu denken. Vergleiche ich mich mit einer Bezugsgruppe, deren Glieder stark aggressiv und dominant, aber wenig hilfsbereit sind, so komme ich vielleicht zu dem Schluß, daß Aggressivität und Dominanz eine relativ geringe Rolle bei mir spielen, während Hilfsbereitschaft mich sehr viel mehr kennzeichnet. Beziehe ich mich aber stattdessen auf eine Gruppe, für welche Hilfsbereitschaft besonders charakteristisch ist, während Aggressivität und Dominanz das Verhalten hier kaum kennzeichnen, gelangte ich vielleicht zu einer Selbsteinschätzung, nach der meine Hilfsbereitschaft nur schwach ausgeprägt ist, während Aggressivität und Dominanz demgegenüber als sehr viel stärker ausgeprägt erscheinen.

Das Problem einer inter-individuell einheitlichen Bezugsnorm besteht bei der Q_s-sort-Methode also nicht anders als bei jeder rating-scale. Zu lösen ist es hier - will man nicht derartige Urteile überhaupt als methodisch unzulässig bewerten - allein mit Hilfe einer entsprechenden Instruktion. So werden die Vpn in diesem Experiment aufgefordert, den Grad der 'intra-individuellen Signifikanz' an jener Verhaltensnorm zu bemessen, welche für Studenten der ersten Semester (nämlich der gegenwärtigen Bezugsgruppe) charakteristisch ist (vgl. dazu die Instruktion im ANHANG). Der Maßstab z.B. für die Beurteilung von Aggressivität wird also nicht durch Definition gesetzt, sondern die Vpn haben ihre Aggressivität im Verhältnis zur (gedachten) 'durchschnittlich üblichen' Aggressivität bei dieser Bezugsgruppe einzuschätzen. Allerdings besteht bei solchem Vorgehen keinerlei Garantie, daß einheitliche Vorstellungen über derartige 'Normen' gegeben sind, noch daß solche Vorstellungen überhaupt vorliegen. Denn 'Norm' in diesem Sinne bedeutet nichts anderes als eine gedankliche Abstraktion von Erfahrungen, Vorstellungen, Erwartungen über ein relativ breit streuendes Verhalten. Jedoch ist auf diesem Wege zumindest ausgeschlossen, daß sehr unterschiedliche Bezugsnormen herangezogen werden, auch wenn der Urteilsmaßstab nur vage vereinheitlicht wird.

Entsprechend konstatiert Cattell (1952) : "...the only instruction that is statistically defensible for Q_s-technique is to ask the subject to rank or rate the traits according

to the extent to which he deviates from the mean therein, i.e., ask how outstanding, in terms of eccentricity, the traits are in his personality- and that is not giving him an easy task"(S.95). Dieser Mittelwert nun, von dem aus die Abweichungen bestimmt werden sollen, ist für Cattell entweder ein individueller für alle Eigenschaften oder derjenige aller Populationen für die betreffende einzelne Eigenschaft. Ein intra-individueller Mittelpunkt für alle Eigenschaften setzte jedoch Normen für jede einzelne Eigenschaft voraus, an welchem diese erst in ihrer Ausprägung gemessen werden können (siehe obiges Beispiel). Es ist daher der Vorschlag von Travers (1951) überzeugender, eine soziale Bezugsgruppe einzuführen, anhand derer zumindest der Extremwert der Urteilsskala definiert werden kann : "The rater might be asked to choose the characteristic on which the individual would have the highest relative standing in his group (which would have to be definded) if all individuals were ranked in order" (S.65). Doch wird im folgenden zu zeigen sein, daß diese Art der Skalen-Definition noch präzisiert werden kann und muß : es besteht nämlich ein wichtiger methodischer Aspekt bei der Q_s-sort-Methode gerade darin, daß nicht d i e Charakteristik frei gewählt werden kann, die einen überhaupt am meisten von der Bezugsgruppe abhebt, sondern daß aus den vorgegebenen Variablen resp. items die r e l a t i v zutreffendste 'Eigenschaft' herausgefunden werden muß.

Eine solche für alle Vpn einheitliche Festlegung der Bezugsgruppe führt zwar dazu, daß die einzelne Vp einen Maßstab für das darzustellende 'solipsistische Kontinuum' erhält, doch stellt sie keine hinreichende Grundlage für den interindividuellen Vergleich dieser introspektiven Urteile dar. Für Cattell z.B. ist der Anwendungsbereich der Q_s-sort-Methode strikt begrenzt auf den intra-individuellen Vergleich; Fragestellungen, die auf der Voraussetzung interindividueller Unterschiede beruhen, müssen ausgeklammert werden : "The very continuum (der 'intra-individuellen Signifikanz', Verf.) is different, because 'significant' is a loose

verbal symbol differently interpreted by different people" (1952, S. 95) [1] [2].

Diese Frage der inter-individuellen Vergleichbarkeit von Q_s-sorts sei zunächst aufgeschoben. Denn weitere Argumente dafür ergeben sich erst, wenn analysiert ist, mit welchem Ertrag zwei 'solipsistische Kontinua' sich intra-individuell vergleichen lassen.

3. Zum Problem der intra-individuellen Vergleichbarkeit solipsistischer Kontinua

Der Vergleich der Einschätzungen einer Vp auf zwei verschiedenen Urteilsebenen, also der Vergleich zweier solipsistischer Kontinua (z.B. von real-self und ideal-self) setzt folgende Momente voraus : das methodologische Moment, daß die Maßdimensionen der beiden Ebenen die gleichen sind, daß sie in äquidistante Skaleneinheiten unterteilt sind und einen gleichartigen Mittelpunkt haben; ferner das rein statistische Moment, jedenfalls bei Anwendung einer Q - K o r r e l a t i o n , daß die zu korrelierenden Werte annähernd normalverteilt und annähernd streuungsgleich sind. Der Forderung zweier annähernd gleicher Normalverteilungen kann in Form der forced-distribution entsprochen werden. Das Problem aber der Vergleichbarkeit der Maßdimensionen (ob intra- oder inter-individuell) ist hier wie bei allen Einstellungs- und

---

1) Auch Stephenson (1953) nimmt diesen Standpunkt - allerdings von ganz anderen Voraussetzungen ausgehend - ein, indem er Fragestellungen dieser Art als für psychologische Theorie von vorneherein irrelevant ansieht. Auf seine Theorie und Methodologie, aus welcher sich diese extreme Konsequenz herleitet, kann hier jedoch nicht ausführlich eingegangen werden.

2) Allerdings räumt Cattell - bei allem zusätzlichen Vorbehalt auch gegen die praktische Realisierbarkeit eines solipsistischen Kontinuums - ein, daß solchen Daten die Charakterisierung als Messung (mensuration) in einem philosophischen Sinne nicht abgesprochen werden solle : "Mathematical processes c a n be carried out on solipsistic measurement, which are real enough, but they must properly be confined to the universe of one mind, which makes their utility extremely limited." (1944, S.298)

rating-Skalen schwieriger zu lösen. Denn es setzt notwendig eine identische Auffassung der Skala bei beiden 'Urteilern' bzw. bei beiden Urteilsvollzügen ein und derselben Vp voraus. Was aber versteht eine Vp unter den Extrempunkten einer Skala, was unter derem Mittelpunkt; wie erlebt sie den Abstand zwischen 2 Skalenpunkten?

Der Vorteil der Q_s-sort-Methode in Form der forced-distribution gegenüber jeder rating scale besteht - wie gleich zu zeigen sein wird - darin, daß diese subjektiven Momente kontrollierbar sind, oder zumindest, daß sie - bei hinreichend kritischer Verwendung dieser Methode (nämlich der Auffassung der Skala nicht als Intervall-, sondern als Ordinalskala, s.u.) - irrelevant sind. Die Vp ist nämlich hier nicht 'frei', z.B. den Extrempunkten der Skala eine subjektiv vorstellungsmäßige Bedeutung zu verleihen; sie ist nicht frei, die Extrempunkte zu umgehen und die Urteile allein anhand des mittleren Skalenbereichs zu formulieren, wie es ebenfalls nicht möglich ist, die Urteile in extrem negative und extrem positive Antworten zu dichotomisieren. Bei der forced distribution ist die Vp vielmehr gezwungen, die ganze Skala zu benutzen und eine vorgeschriebene Anzahl von Differenzierungen (discriminations) vorzunehmen.

Damit wird nicht nur einem Urteils-set entgegengewirkt, mit welchem etwa die Extrempunkte besonders bevorzugt oder aber vermieden würden. Wichtiger ist, daß die Skala selbst nicht mit Hilfe subjektiver Vorstellungen über ein bestimmtes graduell abgestuftes Kontinuum der Ausprägungsgrade einer 'Eigenschaft' zu 'konkretisieren' ist. Das heißt : der Benutzung der Skala geht nicht eine je subjektive 'Definition' durch den Benutzer voraus, wie es der Fall wäre, wenn jede Eigenschaft (jedes item) innerhalb eines eigenen Kontinuums ihrer möglichen Ausprägungsgrade einzuschätzen wäre. Auch wenn die Urteile am Maßstab einer Bezugsgruppe auszurichten wären und dadurch intra-, wie inter-individuell vereinheitlicht würden, hätte der Urteiler doch subjektiv zu entscheiden : "welches Verhalten sehe sich als extrem aggressiv, bzw. inaggressiv oder als durchschnittlich aggressiv an";

dies mit der Folge, daß die Auffassung der 'Bedeutung' der Skalenpunkte in den (durch die Bezugsgruppe) gesteckten Grenzen inter-individuell noch erheblich variieren kann. [1] Gleiche Punktwerte bezeichneten nicht, jedenfalls nicht notwendig, den gleichen Gegenstand, was indessen bei der Verwendung derartiger Intervallskalen vorausgesetzt wird.

Werden stattdessen die inhaltlich heterogenen items eines Q_s-sort's in eine vorgegebene Verteilung sortiert und ist etwa - wie im Falle der hier verwendeten 9-stufigen Skala - festgelegt, daß die Extrempunkte mit je 5 items (der mittlere Punktwert mit 16 items etc.) zu besetzen sind, so wird der 'Bedeutungsradius' dieser Skala festgelegt durch nichts anderes als durch die items selbst. Die Vp urteilt nicht : "ich halte mich für extrem aggressiv, also muß ich allen Aggressions-items den positiven Extremwert zuordnen" (hier+4, s.S.148); sondern die Vp steht vor der Aufgabe, aus dem Gesamt aller Karten jene 5 items auszuwählen, welche - unter den Gegebenen - als "höchst charakteristisch für mich" gelten können. Mit dieser Wahl wird gerade nicht ausgesagt : "diese items treffen genau das, was ich als höchst charakteristisch für mich erlebe", sondern sie besagt nur "v o n d e n g e g e b e n e n items sehe ich diese 5 items als am (relativ) meisten zutreffend an". Das Einsortieren eines items unter den positiven (negativen) Pol der Skala schließt also nicht aus, daß andere (nicht vorgelegte) Charakteristika für dieselbe Vp weit stärker (geringer) zutreffend sind. Die Vp 'definiert' daher nicht die Extrempunkte, indem sie festlegt, welchen Erscheinungen überhaupt das Prädikat 'höchst zutreffend für mich" zuerteilt werden kann, - was u.U. zur Folge hätte, daß keinem der gegebenen items dieses Prädikat zugesprochen würde. Die Vp legt nur die vorgegebenen items in eine Rangreihe (Ordinalskala), bei welcher die Rangstufenzahl vorgegeben, wie festgelegt ist, wieviel rangmäßig gleichwertige Einschätzungen jeweils einer Rangstufe zuzuordnen sind. Die Endpunkte der Skala sind

---

1) Diese Argumentation steht nicht im Gegensatz zu der des vorigen Abschnittes (s.S. 153ff), denn auch dort wurde der Maßstab einer bestimmten sozialen Bezugsgruppe nicht als ein absoluter, sondern ein relativer eingeführt, der weniger die Urteile tatsächlich vereinheitlicht, als die Orientierung an sehr unterschiedlichen Bezugsgruppen verhindert.

'definiert' durch den inhaltlichen Umfang des dargebotenen Materials, welches nicht an einem (subjektiven oder objektiv vorgegebenen) Maßstab zu 'bewerten', sondern nur zu 'ordnen' ist. Es werden nicht in freier Einschätzung den items bestimmte Skalenpunkte zugeteilt, sondern es werden umgekehrt die items - in vorgeschriebener Form - den Skalenpunkten zugeordnet.

Aus diesem Grunde sind auch die Abstände zwischen den Skalenpunkten nicht mehr von der Vp in irgendeiner Weise 'aufzufassen' : der Abstand zwischen 2 benachbarten Skalenpunkten ist vielmehr definiert durch die festgelegte Zahl der items, die von jeder Vp in diese 2 Rubriken einsortiert werden müssen. Zwar ist so nicht eine Äquidistanz der Skalen-Einheiten im streng quantitativen Sinne gewährleistet (wie bei einer Intervall-Skala), doch sind inter-individuelle Unterschiede in der 'Auffassung" der Skaleneinheiten vermieden. Soweit aber die Auffassung der Skala nicht inter-individuell variieren kann, kann die Skala selbst als "experimentell kontrolliert' gelten. Wie sind jedoch die Daten, die eine solche Skala liefert, ihrerseits zu bewerten?

Der absolute Vergleich zweier 'Messungen' setzt eine Intervall-Skala voraus : nur dann könnten Differenzen-scores gebildet werden, welche beispielsweise aussagen würden, daß bei Vp X der Wunsch, aggressiv zu sein, um so und so viele Einheiten größer ist, als ihr 'tatsächliches' Verhalten (so wie sie es erlebt). Hier jedoch sind nur relationale Aussagen möglich (von der Form 'größer-kleiner') und zwar Aussagen, die zugleich nur in Relation zu dem Gesamt der vorgegebenen items bewertet werden können, also in diesem Sinne 'relativ' sind. Das sei in einem Beispiel veranschaulicht :

Wenn eine Vp z.B. im ideal-self nAut am höchsten einschätzt, während dieselbe Variable im real-self nur einen mittleren score erhält, so kann man dem nur entnehmen, daß im Vergleich zu allen anderen Variablen nAut am meisten angestrebt wird, daß ihr aber im Verhalten nicht in vergleichbarer Weise auch das stärkste Gewicht zukommt. Die Vp möchte idealiter in erster Linie (von den gegebenen Variablen) durch Autonomie gekennzeichnet sein, ist realiter aber vorrangig durch andere Variablen (von den gegebenen) charakterisiert. - Oder eine Vp schätzt nAgg im Wunsch-Ich sehr hoch, im ideal-self aber sehr gering ein. Es folgt daraus nicht mehr, als

daß die Vp von allen gegebenen Variablen wunschmäßig in erster Linie Aggressivität realisieren möchte, daß im Ich-Ideal aber gerade diese Variable im Vergleich zu den übrigen besonders negativ bewertet wird.

Sind derartige relationale Aussagen im intra-individuellen Bereich also methodisch durchaus zulässig, so ist jedoch noch nicht entschieden, welche 'Realität' durch sie indiziert wird. Vielmehr wäre erst empirisch zu prüfen, ob dieses Verfahren tatsächlich auf die erlebten Differenzen zwischen 'Wunsch' und 'Verhalten' oder 'Verhalten' und 'Anspruch' etc. verweist, ob es sich um ein angemessenes Korrelat für den Begriff des 'Ebenen-Gefüges' handelt. Da jedoch die 'erlebten Differenzen' nicht als solche realisierbar sind, sondern ein theoretisches Konzept darstellen, das auch seinerseits nur operational repräsentiert werden kann, ist eine Bewährung dieser Methode nur nach dem Prinzip der construct-validity möglich. Nach diesem Prinzip bewährt sich das Verfahren, wenn eine Verknüpfung der Q_s-sort-Daten mit den Daten der WP zu Ergebnissen im Sinne der vorgeordneten theoretischen Erwartungen führt.

4. Zum Problem der inter-individuellen Vergleichbarkeit solipsistischer Kontinua

Würde Cattell (z.B. 1944) gegen solchen intra-individuellen Vergleich introspektiver Daten nichts einwenden, so lehnt er doch den inter-individuellen Vergleich derartigen Materials als methodologisch unvertretbar ab, da solipsistische Daten bestenfalls ipsativ, nicht jedoch normativ (also als Gruppennorm) weiterverarbeitet werden können. Zu prüfen ist daher nun, ob Normativität in diesem strengen Sinne notwendige Voraussetzung für den inter-individuellen Vergleich ist bzw. in welchem modifizierten Sinne auch introspektive Daten 'normativ' betrachtet werden können.

Die Argumentation kann hier diejenige des letzten Abschnittes wieder aufnehmen. Dort war schon eingeschränkt worden, daß aus den Q_s-sort-Daten nicht normativ gefolgert werden könne : 'Vp X erlebt sich als aggressiver als Vp Y sich erlebt'. Denn nicht das Urteil über die eigene Aggressivität

selbst wird (anhand eines absoluten oder normativen Maßstabs) quantifiziert, sondern es wird die 'intra-individuelle Signifikanz' der Aggressivität nur i m V e r h ä l t n i s zu anderen Verhaltens-Charakteristika gewichtet, rangmäßig bewertet. Diese Einschränkung jedoch erlaubt den weiteren Schritt, gerade diesen Grad der 'intra-individuellen Signifikanz', ausgedrückt als eine Verhältnisaussage, inter-individuell zu vergleichen. Es kann aus solchen Daten gefolgert werden : 'im Rahmen der vorgegebenen Variablen gibt Vp X der Variable 'Aggression' ein größeres subjektives Gewicht als es Vp Y tut'. Zwar kann nicht der 'Grad an Aggressivität' inter-individuell verglichen werden, doch lassen sich innerhalb einer bestimmten Gruppe die 'subjektiven Gewichtungen' von Aggressivität aufeinander beziehen. Der Begriff der (Gruppen-)'Norm' basiert hier nicht auf objektiven Verhaltensdaten, sondern bezeichnet ein inter-individuelles Kontinuum verschiedener Grade der 'intra-individuellen Signifikanz'. Damit ist zwar die Einschränkung verknüpft, daß dieses Kontinuum sich nicht in Form einer Intervall-Skala, sondern nur in Form einer Ordinal-Skala quantifizieren läßt, jedoch nicht zugleich auch die Unzulässigkeit inter-individuellen Vergleichens überhaupt.

Die Bedeutung dieser inter-individuellen Differenzen (hinsichtlich der intra-individuellen Signifikanz) kann nur e m p i r i s c h festgestellt werden : eine empirische Validierung jedoch ist im Rahmen dieser Arbeit nicht durchführbar. Wir gehen daher - sozusagen probeweise - davon aus, daß das Verfahren tatsächlich für eben die Merkmale valide sei, für deren Messung es konstruiert wurde. Sodann prüfen wir jene Hypothesen, die sich aus der Annahme einer solchen Validität ergeben. Die Bestätigung dieser Hypothesen kann als Hinweis darauf verstanden werdem, daß das Verfahren zumindest in den Aspekten, auf welche diese Hypothesen ausgerichtet sind, der vorgeordneten 'theoretischen Realität' entspricht, sich also in diesem Sinne 'bewährt'.

Es werden daher nicht auf unzulässige weise introspektive Daten i n t e r -individuell verglichen, als ob es sich um objektive Verhaltensdaten handelte, sondern es wird nur vorläufig und probeweise angenommen, daß den inter-individuellen Differenzen in der subjektiven Gewichtung (soweit sie durch

das verwendete Verfahren hervorgebracht werden) tatsächliche Differenzen des Selbstbildes entsprechen. [1)]

Nicht die 'tatsächliche' Größe dieser Differenzen interessiert hier in einem absoluten Sinne (denn dafür wäre eigens ein Maßstab zu entwickeln). Vielmehr ist mit solcher Voraussetzung einer Validität nicht mehr als folgende Annahme verbunden : je größer die operational feststellbare (inter-individuelle) Differenz zwischen 2 Vpn i.b. auf die von ihnen angegebene intra-individuelle Signifikanz einer Variable ist, je größer ist auch die tatsächliche Differenz in der Selbstwahrnehmung dieser Variablen bei beiden Vpn.

Allerdings würde auch dieses Vorgehen trotz seiner ausdrücklichen Einschränkungen auf relationale, statt im strengen Sinne quantitative Aussagen, nicht von Cattell akzeptiert werden. Es ist also weiter zu prüfen, wie weit es dem bei Cattell vertretenen methodologischen Maßstab genügt, der sich etwa im folgenden ausdrückt : " ... the mathematical psychologist has a right to demand, as a condition of measurement in any reasonable use of the word, (a) that the measurement shall lie in a single continuum, of constant and definable deriction, and (b) that this continuum shall be the same for all subjects." "...it is t h e standard objection to introspective 'measurement' that either or both of these criteria are n e v e r satisfied therin." (1944, S.297)

Gegen diese Cattell'sche Kritik läßt sich anführen, daß in unserem Verfahren die Maß-Dimension nicht - im strengen Sinne - solipsistisch ist. Es handelt sich wohl um introspektive Daten wie um eine ipsative Messung : Doch wird der Maßstab für die Einschätzung der 'intra-individuellen Signifikanz' in gewissem Grade durch eine zusätzliche i n t e r-individuelle Bezugsnorm (d.h. eine einheitliche Bezugsgruppe, vgl.S.154) vereinheitlicht. Ferner ist dieses Verfahren dadurch charakterisiert, daß aus dem Gesamt aller nur denkbaren subjektiven Universa eine solche item-Stichprobe aus-

1) Solches Vorgehen entspricht - ebenso wie es für eine begründete Interpretation der intra-individuellen Differenzen herangezogen wurde (s.S. 160) - dem Prinzip der construct-validity.

gewählt wird, welche inter-individuelle Differenzen notwendig voraussetzt. Es ist ja nicht eine repräsentative item-Stichprobe aus einem bestimmten subjektiven Universum hergeleitet worden, anhand dessen nun eben dieses Subjekt seine eigene Selbstauffassung auf best-mögliche Weise widerspiegeln soll. [1] Vielmehr ist ein allgemeines 'Universum' ganz allgemeiner Variablen konzipiert worden, von welchem angenommen werden kann, daß es grundsätzlich für jedes Individuum relevant ist. [2]

Diese Variablen bzw. items sind allen Vpn gleichartig vorgegeben; jede Vp hat sich anhand des gleichen Instrumentariums darzustellen. Sie wählt beispielsweise nicht aus etwa 500 persönlichkeits-beschreibenden items jene 100 items selbst aus, welche das subjektive Selbstbild in positiver oder negativer Weise am besten repräsentieren. Das 'solipsistische Kontinuum' entsteht vielmehr dadurch, daß die vorgegebenen items so zu 'sortieren' sind, daß das entstandene 'pattern' (wenn auch nur als sehr grobmaschiges Netz) am angemessensten den verlangten Ausschnitt aus dem eigenen Selbstbild deckt : womit vorausgesetzt wird, daß j e d e Vp sich unter dem Blickwinkel eines solchen Ausschnitts von

1) Ähnlich fragen z.B. Crowne und Stephens (1961), ob es nicht in der Konsequenz eines phänomenologischen Ansatzes läge, für jedes Subjekt eine andere, nämlich die ihm allein adäquate Stichprobe zusammenzustellen. Das würde aber bedeuten, daß eine Selbstdarstellung sozusagen nur 'monadisch' möglich wäre, ohne den Vergleichsmaßstab anderer Personen, ohne eine Bezugsgruppe, was aufgrund des sozialen Charakters jeglichen Verhaltens kaum denkbar ist. Weiterhin impliziert eine solche Konsequenz, daß das jeweilig bestimmte Q_S-sort-Verfahren ein Abbild der 'gesamten Persönlichkeit' liefern solle und könne. Darauf wird in einem eigenen Abschnitt über das Problem der Repräsentativität näher einzugehen sein.

2) Diese Sicht ist vergleichbar (wenn auch nicht identisch) mit der Konsequenz Cattells (1952), daß Q-Korrelationen die Ergebnisse der R-Korrelation voraussetzten. Q-Korrelation beinhaltet demnach die Anwendung allgemeiner Persönlichkeitsvariablen ('Faktoren') auf den besonderen Fall eines Individuums. Zum Problem der Allgemeingültigkeit solcher Faktoren siehe z.B. Kaminski 1959, Bergius 1960.

Verhaltenscharakteristika betrachten k a n n .[1]

Allerdings hat eine solche Methode zur Folge, daß die 'wahren' Abstände zwischen der persönlichen Relevanz dieser oder jener Eigenschaft übertragen werden müssen (ähnlich dem geometrischen Begriff der Projektion) auf den vorgegebenen Maßstab eines bestimmten Skalenbereichs, dessen Endpunkte inhaltlich festgelegt sind durch den Bedeutungsradius eben der vorgelegten items, auf einen Maßstab zudem, bei welchem auch die Anzahl der möglichen Differenzierungen zwischen den items festgelegt ist.[2] Das Selbstbild kann nicht einem eigenen Maßstab nach 'frei' formuliert werden, wobei im Einzelfall u.U. mehr oder weniger Differenzierungen vorgenommen werden müßten, sondern seine Darstellung ist (und das gilt gleichermaßen für alle Vpn) in die Grenzen einer bestimmten (und im Prinzip beliebig auswechselbaren) Methode verwiesen. Auf das Problem, daß damit "die zugrundegelegte Methodik ihrer Struktur nach die Realität, auf die sie sich richtet, mitbestimmt" (Kaminski 1959, S.206), wird im nächsten Abschnitt einzugehen sein.

Die Form der forced distribution nämlich, die hier als notwendige Voraussetzung für den intra- und inter-individuellen Vergleich von Q_s-sort-Daten bestimmt wurde, ist in ihren Implikationen näher zu untersuchen : es ist zu präzisieren, welche Information bei solchem Vorgehen gewonnen wird, welche verloren geht.

---

1) Von solchen Voraussetzungen aus ist Stephenson (1953) zu kritisieren, der zum Ziel seiner Theorie 'the total person-in-action' erhebt und damit die Frage interindividueller Differenzen vollständig auszuklammern zu können glaubt. Vgl. Fußnote 1) der Vorseite.

2) "'Wahre' Abstände" soll hier nicht bedeuten, daß es diese in einem absoluten Sinne, also unabhängig von ihrer Feststellung 'gäbe'. Es soll nur deutlich gemacht werden, daß die 'erlebten' Abstände methodisch 'gefiltert' werden, daß z.B. items, zwischen denen kein Abstand erlebt wird, rangmäßig unterschieden werden müssen oder daß ein (erlebter) 'großer' Abstand anhand der Anzahl vorgegebener Rangplätze nicht dem Erleben gemäß wiedergegeben werden kann.

5. Zum Problem der forced-method

Das Problem der Verwendung der forced distribution besteht vor allem darin, daß dieser von außen gesetzte einheitliche Maßstab in sehr unterschiedlichem Maße dem individuellen Selbstbild korrespondiert. Das experimentelle Abbild der angezielten Realität der 'Selbstwahrnehmung i.b. auf bestimmte Eigenschaften' ist - wie bereits oben beschrieben - nicht mehr als die 'Projektion' von erlebten psychischen Gegebenheiten sehr verschiedener Ausdehnung und Struktur auf den engen Maßstab der vorgelegten Skala und der vorgegebenen items. Die vorgeschriebene Form der (angenäherten) Normalverteilung kann in sehr unterschiedlichem Maße der 'tatsächlichen' Verteilung entsprechen (vgl. Fußnote 2 der Vorseite). Denn würde mit 'freier Verteilung' gearbeitet, so könnte sich für einzelne Vpn der Kurvengipfel beim Sortieren derselben items zum positiven oder negativen Ende der Skala verschieben, oder es könnten auch zweigipflige Verteilungen entstehen.

Diese bei der forced-method verdeckten inter-individuellen Unterschiede in der Form der Verteilung können von diagnostisch nicht geringem Interesse sein. Sie verweisen u.a. auf die unterschiedliche Relevanz, welche die vorgegebenen items für verschiedene Vpn besitzen : für eine Vp enthalten die items gerade jene Inhalte, die sie selbst als am wichtigsten für eine Charakterisierung ansieht; eine andere Vp dagegen kann sich sehr viel schwieriger vermittels gerade dieser Persönlichkeits-Beschreibungen darstellen, da sie sich vorrangig in ganz anderen Kategorien erfährt. Die (hier nicht eingelöste) Forderung, daß bei Verwendung der forced-distribution zunächst experimentell zu prüfen ist, ob ein freies Sortieren bei einer Kontrollgruppe zu einer angenähert gleichen Verteilung führt, gewinnt also an Gewicht. [1)]

1) In unserem Experiment ergab sich, daß alle Vpn (trotz einer nur auf 'forced distribution' ausgerichteten Instruktion, siehe ANHANG) zunächst die Karten (items) 'frei' sortierten, um sie dann erst in die geforderte Verteilung umzusortieren. In fast allen Fällen ergab sich eine unimodale, doch selten eine symmetrische Verteilung, in einzelnen Fällen auch eine bimodale Verteilung. Insbesondere aber ergaben sich unterschiedliche Schwierigkeiten bei der Aufgabe, die geforderte Verteilung im nachhinein zu realisieren. Die ersten spontanen Verteilungen konnten nicht mehr systematisch mit ausgewertet werden.

So hat etwa Jones (1956) festgestellt, daß bei f r e i e r Verteilung der Unterschied zwischen den (jeweils zusammengefaßten) Gruppenkurven von Studenten und Psychotikern signifikant war; er fand dabei auch, daß alle Vpn signifikant von der Normalverteilung abwichen und daß die Unterschiede der Verteilungen selbst innerhalb einer Gruppe (Studenten oder Psychotiker) zu groß waren, als daß man jede Gruppe als durch einen bestimmten (durchschnittlichen) Verteilungstyp repräsentiert hätte ansehen können. Trotzdem muß daraus nicht gefolgert werden, daß die Methode "may be psychologically indefensible, if there is a reason to think that persons differ in their variability over traits" (Cronbach 1953, S.379).

Es sind bisher nämlich nicht die Bedingungen geklärt, welche zu der einen oder anderen Verteilungsform führen : es wird dies nicht nur an der Gruppe, sondern ebenso an der jeweils verwendeten item-Stichprobe, wie an subjektiven response-sets liegen. Aus diesem Umstand folgert Block (1961), daß man diejenige Information, die auf dem Wege der forced-method im Vergleich zur freien Verteilung verloren geht, durch bestimmte inhaltlich gezielte items, wie aber auch durch eine hinreichende inhaltliche Breite der items wieder einholen können müsse. Es sind dies ähnliche Überlegungen, wie die hier bereits ausgeführten (s.S.146) : daß nämlich nur eine inhaltlich hinreichende Vielfalt der items eine gewisse Garantie dafür bietet, daß inter-individuelle Unterschiede tatsächlich ihren Niederschlag finden können und dafür, daß eine Normalverteilung überhaupt hergestellt werden kann. Block argumentiert gegen ein unforced-sorting insbesondere damit, daß hier sehr viel weniger 'Unterscheidungen' (discriminations) von der Vp gefordert werden (vgl.S. 157) [1]. Da umgekehrt eine rechteckige Verteilung es ist, welche die meisten Unterscheidungen verlangt (Livson, Nichols 1956), schlägt Block einen Kompromiß vor, der gleichzeitig hinreichend viele 'discriminations', wie die rechnerischen Vorteile der Normalver-

1) Je weniger Skalenpunkte von der Vp benutzt und je mehr items damit einzelnen Skalenpunkten zugeordnet werden, desto weniger muß die Vp zwischen den einzelnen items 'unterscheiden'.

teilung mit einschließt : nämlich eine relativ abgeflachte Normalverteilung.

Dieser Block'schen Argumentation schließt sich auch unsere Arbeit an : Im mittleren Bereich, der psychologisch am wenigsten informativ und in welchem eine genaue Differenzierung der Einschätzungen am schwersten zu leisten ist, werden am wenigsten 'Unterscheidungen' gefordert (viele items pro Skaleneinheit); in den Extrembereichen dagegen, welche am wichtigsten für eine Charakterisierung sind, werden die meisten discriminations verlangt (nämlich pro Skaleneinheit relativ wenig items), was insbesondere vorteilhaft deshalb ist, weil auch der Korrelationskoeffizient gerade auf diesen Extrembereichen am stärksten beruht. [1)]

Welche Form der Verteilung zu wählen ist, hängt jedoch nicht allein von methodologischen Überlegungen ab, sondern ebenso von der Aufgabenstellung (vgl. Fußnote 1) . Für die Zwecke dieser Arbeit muß ein Verfahren geschaffen werden, welches - auf Kosten welcher Information auch immer - bei introspektiven Daten intra- und inter-individuelle Vergleiche zuläßt. Nur die forced-distribution erfüllt diese Forderung, wenn auch in dem eingeschränkten Sinne, daß nur Vergleiche relationaler Art möglich sind. Welche Information nun den forced-distributions verschiedener Vpn entnommen werden kann, sei noch einmal an einem Beispiel ins Gedächtnis gerufen :

Diejenigen Vpn, welche die items der Variable nAgg im Gegensatz zu anderen Vpn der gleichen Versuchsgruppe besonders hoch einschätzen, sprechen, am Maßstab der übrigen vorgegebenen Variablen, in relativ ähnlicher Weise gerade nAgg eine besondere subjektive Relevanz zu; sie unterscheiden sich gemeinsam von jenen Vpn, welche mit nAgg demgegenüber eine besonders geringe intra-individuelle Signifikanz verbinden.

welcher tatsächliche Unterschied an offener Aggression diese beiden Untergruppen charakterisiert, ist zwar - wie

---

1) Dennoch wird hier nicht mit Block gefolgert, daß die forced method von vorneherein wegen ihrer methodischen Vorteile auch für jeden Zweck das angemessenste Verfahren ist. Da die forced method die je individuelle Eigenart des freien Sortierens unterschlägt, kann eine zusätzliche freie Verteilung wichtige diagnostische Hinweise liefern. Da umgekehrt die freie Verteilung Ausdruck von response-sets verschiedener Art sein kann, ist die forced method hier ein wichtiges Korrektiv. Allerdings ist die Bedeutung verschiedener (freier) Verteilungsformen bis heute nicht untersucht worden.

nun mehrfach betont – aus dem Verfahren selbst nicht ablesbar. Doch kann man - wiederum probeweise - annehmen, daß diese Vpn ihr Verhalten tatsächlich (in derselben Richtung) unterschiedlich e r l e b e n . Jedenfalls wird der Unterschied an Aggressivität zwischen den Gruppen größer sein, als derjenige zwischen den Gliedern einer dieser Untergruppen. Dieses aber ist - bei aller nur denkbaren Vielfalt an 'psychischer Realität' - gerade diejenige Information, welche für die Zwecke des hier vorgenommenen Experiments vorrangig interessiert.

Die vorläufige Annahme, daß das Verfahren im intendierten Sinne valide sei, läßt sich - wie ebenfalls schon eingehender behandelt - allein durch die A n w e n d u n g der Methode bestätigen oder verwerfen. Allein hier kann sich zeigen, ob die inter-individuellen oder auch Gruppen-Unterschiede, die das Verfahren hervorbringt, sich innerhalb jener Hypothesen bewähren, auf welche hin dieses Verfahren konstruiert wurde. Die Anwendung der forced method macht also eine Gültigkeitsprüfung notwendig (in Form der construct-validity), auch wenn die forced method sich im Rahmen einer bestimmten Aufgabenstellung als methodisch besonders angemessen anbietet. Sollte sich umgekehrt die forced method als d i e angemessene Methode aus einem theoretischen Konzept vollständig deduzieren lassen, so müßte vorausgesetzt werden können, daß der zu messende Gegenstand selbst schon 'normalverteilt' ist.

Daß diese Voraussetzung der Normalverteiltheit der Eigenschaften bei allen Verfechtern der forced method implizit gegeben sei, ist durchaus behauptet worden (vgl. Livson, Nichols 1956); Jones (1956) spricht sie explizit aus. Gehen also psychologische Voraussetzungen in diese Methode ein, die theoretisch nicht vertretbar sind? So fragt auch Kaminski : "Muß man damit nicht zugleich alle dem mathematischen Modell der Normalverteilung inhärenten Grundannahmen mit übernehmen, ohne daß diese vielleicht zu unseren sonstigen psychologischen Vorstellungen über Urteilsgeschehen und Bezugssysteme passen?" (1959, S. 220).

Es ist nun hier wieder an die Diskussion über das Verhältnis von 'Theorie' und 'Experiment' bzw. von 'theoretischem' und 'experimentellem' Satz anzuknüpfen (vgl.S.99ff). Eine

Methode ist nicht unter dem Gesichtspunkt ihrer (gewollten oder ungewollten) inhärenten V o r a u s s e t z u n g e n , nicht allein von den Realitätsbestimmungen her, die ihr offenbar anhaften, zu betrachten : das Verfahren selbst ist nicht die Definition der zu messenden Realität. Eine Methode ist vielmehr unter dem Aspekt ihrer K o n s e q u e n z e n zu analysieren : ihrer Rolle nämlich, die sie für die Art der Abbildung der gemeinten 'theoretischen Realität', für deren 'Repräsentation' spielt. Es sind die M e ß e r g e b n i s s e , welche durch die Bedingungen ihres Zustandekommens bestimmt (und begrenzt) sind; nicht ist es die 'theoretische Realität', über welche durch das Verfahren entschieden würde. Zu analysieren ist daher bei jeder Methode, ob sie die theoretische Realität, um welche es in der betreffenden Fragestellung geht, hinreichend 'abbildet'. So muß es keinen Widerspruch bedeuten, wenn anhand der forced method Daten gewonnen werden, die 'repräsentativ' für eine bestimmte 'theoretische Realität' sind, welche ihrerseits unabhängig von der Frage der Normalverteiltheit der betreffenden 'Eigenschaften' definiert werden kann. Die Methode ist also nach der Natur ihrer Ergebnisse zu beurteilen (nämlich nach dem Grad ihres 'Herstellungsmoments', (vgl. Holzkamp, a.a.O.) und nicht nach der Natur der durch sie vermeintlich vorausgesetzten psychischen Realität.

Die 'gemeinte' psychische Realität nämlich wird bereits schon durch jenes psychologische Modell definiert, innerhalb dessen eine bestimmte Methode - nur als ein Werkzeug hier fungierend - überhaupt erst ihre Anwendung findet. Es verlagert sich somit die Frage nach der Kongruenz von Methode und 'psychischer Realität' auf die Frage der Angemessenheit einer bestimmten Methode innerhalb eines bestimmten Modells. Das Modell ist es, welches ein bestimmtes gedankliches Abbild der 'psychischen Realität' voraussetzt : die Maßmethode hat sich innerhalb dieses 'Abbildes' nur zu bewähren. An sich selbst ist die Methode nicht mehr als ein indifferentes Werkzeug : ihre Rolle wird erst definiert im Rahmen der Theorie. (Siehe dazu auch Hörmann 1964b).

Das Verfahren der forced distribution ist also wie jede nur denkbare Methode in der diagnostischen und experimentellen Psychologie begründbar und auch kritisierbar nur im Rahmen der Theorie, in der es zur Anwendung kommt.

Schließlich und letztens ist in Hinsicht auf die Methode der forced distribution nun noch einem besonderen kritischen Punkt Aufmerksamkeit zuzuwenden : dies vom Gesichtspunkt der Statistik aus, welche - jenseits aller allgemeinen Fragen der Methodologie - ihre eigenen mathematischen Beurteilungskriterien hat.

Dafür ist noch einmal kurz ins Gedächtnis zu rufen, wie hier die Reaktion auf die Q-sort-Skala charakterisiert wurde. Als Folge der forced method ist aufgezeigt worden, daß der Abstand zwischen 2 Skalenpunkten nicht in einem direkt quantitativen Sinne zu verstehen ist, sondern daß dieser Abstand für alle Vpn gleichartig definiert ist durch die jeweils festgelegte Anzahl der items, die auf 2 benachbarte Skalenpunkte einzusortieren sind (s.S.).Dadurch ist auf der einen Seite gewährleistet, daß für den Pb nicht etwa die Möglichkeit gegeben ist, den Abstand zwischen 2 Skalenpunkten so oder so aufzufassen : er hat vielmehr zu entscheiden, welche der items er als relativ gleichwertig beurteilen kann und welche relativ mehr oder weniger relevant sind. Das hat aber auf der anderen Seite zu Folge, daß es sich bei der Q_s-sort-Skala in Wahrheit nicht um eine Intervall-, sondern um eine Ordinal-Skala handelt : anders ist die geforderte Reaktion bei einer forced method psychologisch jedenfalls nicht zu sehen.

Es taucht damit das Problem einer eigenartigen Vermischung auf, das in der Literatur bis heute offenbar nicht gesehen worden ist und allerdings auch in dieser Arbeit ungelöst stehen bleiben muß. Psychologisch gesehen, also vom Reaktionstyp her beurteilt, handelt es sich um eine Ordinalskala; statistisch gesehen, nämlich von der Verarbeitung der Daten aus betrachtet, wird so verfahren, als ob es sich um eine Intervallskala handelte. Dieses gilt zwar nicht in Hinsicht auf die hier vorgenommene Diskussion möglicher interindividueller Vergleiche, so weit sie den beschriebenen

relationalen Charakter haben, also an den Bedingungen der Ordinalskala gerade ausgerichtet sind. Es gilt dies aber für die Verwendung der Produkt-Moment-Korrelation, wie sie weithin und auch in dieser Arbeit vorgenommen wird. Die Vermischung besteht nämlich darin, daß man hier künstlich eine Normalverteilung auf der Basis einer Ordinalskala aufgestellt hat.

Man kann hier - erschwerenderweise - nicht argumentieren, daß bei freier Verteilung ein Q_s-sort nachweislich angenähert einer Normalverteilung entsprechen kann. Es handelte sich nämlich bei einer freien Verteilung gerade um eine Intervallskala, indem nämlich hier die Punktwerte der Skala metrisch und nicht relational aufgefaßt werden würden : die Vp müßte hier jedem Punktwert eine bestimmte eigene Bedeutung zuschreiben und hätte ferner die Instruktion erhalten, die Abstände zwischen den Skalenpunkten als gleich aufzufassen und sie somit in einem quantitativen Sinne zu verstehen. Nur für den Sonderfall, daß bei einer freien Verteilung sich 'von selbst' eine Normalverteilung einstellte, wäre - einem strengen statistischen Maßstab nach - gerechtfertigt, die Produkt-Moment-Korrelation vorzunehmen.

Ein weiterer statistischer Gesichtspunkt noch ist, daß die Einschätzungen der items nicht unabhängig voneinander vorgenommen werden können. Sundland (1962) hat darauf hingewiesen, daß die (bei einer hinreichenden Vielzahl von Persönlichkeitsbeschreibungen) unvermeidlichen inhärenten Interrelationen zwischen den items zu einer 'spurious correlation', nämlich einer 'Inflation' der Koeffizienten führe. Er folgert, der einzige Ausweg (possible hope) "would be the use of items whose interrelatedness is known." Eine solche Forderung jedoch hat hier nicht realisiert werden können. Andere Autoren versuchen den strengen statistischen Maßstab durch einen einfacher zu realisierenden, nämlich einen psychologischen Maßstab zu ersetzen. So sagt etwa Stephenson : "Each item of the sample has to be appraised as a dinstinct and separate act or operation on X's (the subject's, Verf.) part". (1953, S. 65) Ein etwas klareres Kriterium liefert

Block mit dem Begriff der 'conceptual independence', unter der verstanden wird "that the psychological sense of each item could not be coordinated to or derived from the psychological sense of any other item or conjunction of items." (1961, S.53) Doch handelt es sich bei diesen Versuchen eher um einzelne und durchaus wichtige Kriterien für die item-K o n s t r u k t i o n ; die statistische Forderung aber nach gegenseitiger Unabhängigkeit der E i n s c h ä t z u n g e n selbst ist weder durch sie, noch offenbar überhaupt im Rahmen solipsistischer Kontinua, die auf einer forced method beruhen, zu realisieren. Es ist dies also eine Frage, die allgemein heute noch nicht erschöpfend geklärt worden ist. Da es aber gerade die Anwendung der Korrelation ist, die eine forced method voraussetzt, und da es wiederum die forced method ist, die - wegen der nicht bestehenden Garantie unabhängiger Einschätzungen - zu spurious correlation führt, kann hier nur mit entsprechender Vorsicht eine solche Produkt-Moment-Korrelation vorgenommen werden. Es wird im Kap. VII zu zeigen sein, daß - bei entsprechender Zusammenfassung der Daten - auch die Rangkorrelation hier anwendbar ist : als nonparametrisches Verfahren hat sie zwar nicht so strenge Maßstäbe; doch basiert sie auf sehr viel weniger Daten und ist so aus diesem Grunde nicht minder ungenau, auch wenn sie statistisch gesehen das angemessenere Verfahren darstellt.

Sind nun bisher die Probleme besprochen worden, welche insgesamt auf die Anwendung der forced distribution zurückgehen, so sind im folgenden noch weitere Probleme zu behandeln, welche mit der Q_s-sort-Methode als solcher verknüpft sind, unabhängig von der besonderen Form einer festgelegten Verteilung.

6. Zum Problem der Repräsentativität der item-Stichprobe

Die Frage nach der Repräsentativität der item-Stichprobe nimmt in der Literatur eine beherrschende Rolle ein. Ihr liegt ein Denkansatz zugrunde, der näher zu bestimmen ist, soll diese Frage sich für das hier verwendete Q_s-sort beantworten lassen.

Bemerkenswert ist nämlich, daß dieses Problem bei den übrigen self-report-Techniken nicht in den Blick geraten ist, obwohl doch auch diese auf einer item-Selektion beruhen. Beim MMPI etwa sind die items nicht als Stichprobe jenes Universums aufgestellt worden, welches die gesamte zu messende psychologische Dimension (z.B. Paranoia) ausmacht : man hat vielmehr die Gesamtheit der items aus den Definitionen der angezielten psychischen Dimensionen hergeleitet, und dann nach dem Prinzip der 'concurrent validity' (Cronbach, Meehl 1955) ihre Gültigkeit an entsprechenden pathologischen Gruppen in Gegenüberstellung zu einer 'normalen' Population untersucht und das Verfahren daran geeicht.

Daß im Rahmen der Q_s-sort-Methode die Frage der Repräsentativität der items zu solchem Interesse gelangte, liegt offenbar daran, daß man ursprünglich in noch sehr unkritischer Weise glaubte, auf diesem Wege d a s Selbstbild als solches erheben zu können : die Frage nach der Repräsentativität ist also die erste kritische Einschränkung solch vorschneller Generalisierung. So sieht etwa noch Mowrer (1953) den besonderen Vorzug der Q_s-sort-Methode (bei ihm S-sort) darin, daß sie auf Grund ihres 'atomism' (nämlich auf Grund genügend vieler, wie genügend kleiner Informationseinheiten) einen großen Schritt weiterführe auf dem Wege der Messung der 'whole person'. Diese Annahme wird freilich empirisch durch die sehr unterschiedlichen Ergebnisse bei Verwendung verschiedener item-Zusammenstellungen eingeschränkt: "...we may note that while using an S-sort it may look as if we are studying 'the whole personality' but we are actually not, since we do not have a random sample of items drawn from all possible items relevant to any given individual. ...the whole matter of item selection where trait universes are involved is as yet very incompleted explored and will profit from further inquiry". (S.361). Jedoch wird nicht explizit ausgeschlossen, daß bei genügender Klärung dieser Probleme der item-Selektion die Q_s-sort-Methode direkt als Meßinstrument der 'Persönlichkeit' fungieren könne: "If we were sure that the items of an S-sort were suffiently numerous and r e p r e s e n t a t i v e and if we were sure that

our subjects sorted them with the proper set or attitude, we could be assured of getting for each person a combination of ratings or values which would be unique ..." (S. 348, Sperrung durch Verf.).

Dagegen ist jedoch einzuwenden, daß die Repräsentativität einer solchen Stichprobe, welche den 'ganzen Menschen' (oder in einem ganzheitlichen Sinne einen 'Persönlichkeitsbereich') bestimmen soll, überhaupt nur zu erlangen ist, wenn die Stichprobe aus dem jeweils individuellen Universum dieses bestimmten Menschen gewonnen wurde. Ähnlich argumentieren Crowne und Stephens (1961), wenn sie es als eine Konsequenz des strikt phänomenologischen Ansatzes bezeichnen, daß hier nur für jedes Subjekt die ihm allein adäquate Stichprobe zusammengestellt werden könnte. [1] Entsprechend auch verneint Lowe (1961) eine Vergleichbarkeit und Austauschbarkeit verschiedener Q_s-sorts, also solchen, die auf verschiedenen item-Selektionen basieren, auf Grund der weiteren Überlegung, daß es sich jeweils um operational verschieden definierte Dimensionen handele. Ähnlich kritisieren Crowne und Stephens (1961) das von Butler und Haigh (1954) konstruierte Q_s-sort und den Index of Adjustment and Values von Bills (1951) als zufällige Stichproben (nicht 'Zufallsstichprobe'!), die zugunsten des 'client-centered counseling' zusammengestellt seien und somit über diese spezifischen Inhalte hinaus nicht generalisiert werden könnten. Sie fordern demgegenüber, daß das 'bias' des Autors nicht 'zufällig' - nämlich unkontrolliert und daher unvermeidbar - die item-Konstruktion lenken solle; die items sind vielmehr so zu konstruieren, daß sie als repräsentativ gelten können für die verwendete Theorie und als inhaltlich angemessen für die untersuchte Population.

Hier also wird die Frage nach der 'Repräsentativität' für ein 'imaginäres' und offenbar als faktisch bestehend angenommenes Universum zur Frage nach der Repräsentanz der items

---

1) Innerhalb des rein statistischen Denkmodells ist aber auch diese Einschränkung nicht vertretbar. Statistik ist nur anwendbar auf 'random selecting', nicht auf 'expert selecting' (Wallis,Roberts 1962). Eine Zufallsstichprobe aber ist nur aus einer Population von Individuen zu entnehmen; items dagegen bilden keine Population, aus welcher Zufallsstichproben gewonnen werden könnten (vgl. dazu insbesondere Loevinger 1965, siehe auch Fußnote 1), S.211).

für eine bestimmte vorgeordnete Theorie [1]. Hier werden die items nicht mehr 'selegiert', sondern 'konstruiert'. Es geht nicht mehr um ein Abbild der 'ganzen Person', wie es beispielsweise auch dort implizit intendiert wird, wo der discrepancy-score zwischen real- und ideal-self als Index fungieren soll für den allgemeinen Begriff der 'Anpassung' [2].

Das Weiterführende an dem Ansatz von Crowne und Stephens besteht darin, daß die Begriffe, welche durch das Q_s-sort 'realisiert' werden sollen, zu definieren sind sowohl durch ihre Relation zu den übrigen Konstrukten der Theorie, wie auch durch die verhaltensmäßigen Korrelate : "...the behavioral referents and the hypothesized relationships of the construct are described as parts of its definition" (S.109). Nur so kann vermieden werden, daß jeder Begriff durch beliebige operationale Definitionen ersetzt wird, die dann tatsächlich Verschiedenes messen würden; nur so kann auch vermieden werden, daß die Begriffe auf ihr operationales Korrelat reduziert werden. Vielmehr können die Ergebnisse eines so konstruierten Verfahrens im Sinne der Logik der construct-Validierung geprüft werden, nämlich innerhalb des Netzes von Hypothesen über die verwendeten constructe.

---

1) Daß der Gedanke, die items eines Q_s-sorts müßten sich als 'Zufallsstichprobe' eines bestimmten item-Universums ausweisen lassen, zu theoretisch absurden Konsequenzen führte, sei an einem Beispiel verdeutlicht. Das Problem besteht ja darin, begrenzte Klassenbestimmungen zu finden, nach welchen ein bestimmtes Universum festgelegt werden kann, aus welchem sich seinerseits 'Zufallsstichproben' entnehmen lassen. Hilden (1958) legt als Universum alle Wörter aus dem Thorndike-Lorge'schen Wörterbuch fest, welche sich als statements über menschliches Verhalten, Fühlen, Denken auffassen lassen und zugleich dieselbe Auftretenswahrscheinlichkeit besitzen. Was aber besagt - ohne jeden theoretischen Zusammenhang - eine Stichprobe aus diesem 'Universum', für welchen psychologischen Tatbestand kann sie als repräsentativ gelten?

2) Diese Annahme, daß die 'Anpassung' schlechthin bemessen werden könne, war ja der Ausgangspunkt für die in der Frage nach der Repräsentativität der items ausgesprochene Kritik; doch sollte mit dem Prinzip der Repräsentativität nicht der Anspruch, die 'ganze Person' zu messen, aufgehoben,sondern nur zureichend begründet werden, was seinerseits als methodisch undurchführbar bewertet werden muß.

Einen vergleichbaren Ansatz haben auch Cattell einerseits und Stephenson andererseits. In der Konsequenz seiner Persönlichkeitstheorie fordert Cattell (wie bereits erwähnt) : "to use tests sampled from the principle personality factors already found by R-technique" (1952, S.100). Auch hier sind also die items aus einer Theorie zu deduzieren; allerdings wird das 'sampling' selbst und seine Validität nicht näher beschrieben.

Theoretisch am besten durchgeführt ist der Ansatz von Stephenson (1953), der nämlich, in Anlehnung an die Methodologie von Fisher, die item-Selektion auf eine small-sample-Theorie gründet : "Structuring a sample consists of composing it artificially, instead of selecting it at random from a parent-universe" (S. 66). Wesentlich ist also, auch hier, daß nicht die items als Stichprobe das zugehörige Universum repräsentieren, sondern daß eine bestimmte Strukturierung des sample's durch den Experimentator intendiert und kontrolliert wird. Die Kontrolle beruht hier allerdings auf rein statistischen Voraussetzungen , wie die small-sample-Theorie selbst [1].

Das Konzept der construct-Validität, als Rahmen für eine begründete Anwendung der Q_s-sort-Methode (im Sinne von Crowne und Stephens, s. o.) ist demgegenüber für den Zweck dieser Arbeit das angemessenere, da das Q_s-sort hier nicht selbst schon, wie bei Stephenson, die methodische Transformation einer Theorie oder bestimmter ihrer Propositionen darstellt, sondern in seiner Zuordnung zu anderen Verfahren bestimmte empirische Information liefern soll, deren Gültigkeit zumindest indirekt (also nicht unbedingt an einem Außenkriterium) festzustellen bleibt und nicht vorausgesetzt werden kann.

Zusammenfassend läßt sich feststellen, daß die Frage nach

---

1) Da es sich jedoch hier zugleich um eine eigene Methodologie handelt, nach welcher es weniger um den methodischen Zugang zu bestimmten empirischen Problemen geht, als vielmehr um den Beweis deduzierter Propositionen und damit um den Beweis einer Theorie allein vermittels der Q-technique, sei hier nicht näher auf diesen Ansatz eingegangen.

der Repräsentativität der item-Stichprobe fehlleitet : die items eines Q_s-sort's lassen sich nicht als eine 'Zufallsstichprobe' eines 'Universums' auffassen, für welches das Verfahren dann 'valide' wäre, wenn die A u s w a h l - Methoden Repräsentativität garantierten. Vielmehr können die items nur als 'repräsentativ' für eine bestimmte Theorie gelten, aufgrund von deren Annahmen sie als 'experimentelle Realität' hergestellt, nämlich k o n s t r u i e r t werden. Ihre Gültigkeit kann nicht mehr vorausgesetzt werden, sondern ist im Rahmen der zugehörigen Theorie zu bestimmen.

Die Validitätsprüfung führt hier allerdings zu ähnlichen Problemen, wie sie für die projektiven Verfahren mit dem Begriff der 'differentiellen Validität' bereits dargelegt wurden. Auch sie werden jedoch bis heute eher als Widersprüche, die den Verfahren anhaften, diskutiert, denn präzise auf Populationsdifferenzen zurückgeführt. Diese Populationsdifferenzen bestehen zumindest - systematische Analysen gibt es bis heute nicht - darin, daß entweder mit oder ohne Verteidigungsverhalten reagiert wird.[1)]

---

1) Dies zeigt sich in Untersuchungen, die eine kurvilineare Beziehung zwischen dem discrepancy-score (zw. real- und ideal-self) und einem Anpassungsindex nachweisen : nicht nur ein hoher discrepancy-score korreliert mit schlechter Anpassung, sondern auch ein geringer discrepancy-score, was auf Ich-Verteidigung schließen läßt (vgl. Block,Thomas 1955). Butler und Haigh (1954) nehmen stattdessen eine lineare Beziehung an und schränken sie dadurch ein, daß ein geringer discrepancy-score auch Verteidigungsverhalten indizieren k ö n n e . Wenn jedoch ein Verfahren bestimmte Gruppen innerhalb seiner eigenen Aussagemöglichkeiten nicht differenziert, kann es nur differentiell für diese Gruppen validiert werden; denn es handelt sich nicht um eine einheitliche Population und entsprechend bei den discrepancy-scores nicht um ein einheitliches Kontinuum.
Andere Untersuchungen, welche ein Außenkriterium heranziehen, erhalten allerdings direkte Relationen zwischen discrepancy-score und beobachtbarem Verhalten (z.B. 'Aktivität im Studium', 'Glück in der Ehe', etc; vgl. Holt 1951; Chase 1957, Eastman 1958, Turner, Vanderlippe 1958). Untersuchungen, die mit der Wort-Assoziations-Methode als Index für Konflikt arbeiten, haben unterschiedliche Ergebnisse (vgl. Roberts 1952, Bills 1953, Zimmer 1954). Brownfain (1952) legt seiner Untersuchung die Differenz zwischen 'positivem' und 'negativem' Selbstbild zugrunde und findet eine direkte Relation zu einem Anpassungsindex. Für eine Zusammenfassung von Validitätsuntersuchungen verschiedener Art siehe Lowe (1961).

In dieser Arbeit wird - darauf war schon hingewiesen worden - von dieser Variablen abgesehen, in der Annahme, daß eine entsprechende Motivation (vermittels situativer Determinanten) ein bewußtes Verteidigungsverhalten unwirksam machen könne (vgl. S.135f). Demzufolge wird angenommen, daß Vpn mit durchgehend hoher Übereinstimmung zwischen 2 Bezugssystemen (z.B. zwischen real- und ideal-self) eine (in bezug auf dieses Kriterium) relativ homogene Gruppe bilden und nicht sich dadurch unterscheiden, daß bei ihnen entweder tatsächlich eine derartige Korrespondenz vorliegt oder aber, daß Verteidigungsverhalten gerade die fehlende Übereinstimmung verdeckt.

Die Vorgehensweise dieser Arbeit läßt sich nun dahingehend charakterisieren, daß auch hier die Kriterien, nach welchen die items zu konstruieren sind, allein aus dem besonderen theoretischen Ansatz hergeleitet werden und auf diesen beschränkt sind. Diese Kriterien sind, darauf sei noch einmal verwiesen, im ANHANG aufgeführt. - Die Variablen, welche dem Experiment zugrundeliegen, sind (wie bekannt) dem Murray'schen need-System entnommen : sowohl die items des Q_s-sort's, wie die Auswertungsgesichtspunkte für die WP (s.S. 316) halten sich an die Murray'schen Definitionen, die sich für die Zwecke dieser Arbeit dadurch auszeichnen, daß sie in Form von V e r h a l t e n s -Beschreibungen gegeben sind. - Die items sind insbesondere inhaltlich auf solche Verhaltenssituationen hin ausgerichtet worden, welche für Studenten charakteristisch sind und nicht in eben dem Maße auch für andere Gruppen, so daß eine Generalisation über diesen Rahmen hinaus von vorneherein ausgeschlossen wird [1]. Die vorangestellte Theorie enthält Annahmen über Relationen zwischen den Daten des Q_s-sort-Verfahrens und der WP : ihre Bestätigung impliziert zugleich - innerhalb der Theorie - eine Bewährung des Q_s-sort-Verfahrens selbst.

1) vgl. dazu die bereits erwähnte Forderung von Crowne und Stephens (1961), s.S.175.

Weitere Annahmen, die sich aus den Begriffen 'real-self', 'ideal-self' und 'Wunsch-Ich' deduzieren lassen und nach dem Prinzip der construct-Validität als Kriterium dafür dienen können, wieweit die 'theoretische Realität' tatsächlich durch das Q_s-sort-Verfahren repräsentiert wird, sind in den folgenden Abschnitten herzuleiten.

7. Zum Problem der Variable 'social desirability'

Die Diskussion der Variable 'social desirability' nimmt in der Literatur einen besonders großen Raum ein [1]. Eine Reihe von Autoren glauben nachgewiesen zu haben, daß die items eines self-reports weniger danach von dem Pb beurteilt werden, wie sie ihn 'tatsächlich' charakterisieren, als danach, ob sie allgemein 'erstrebenswert' oder doch jedenfalls 'akzeptabel' sind. Es würde dies bedeuten, daß man auf dem Wege von self-reports weniger das tatsächliche Selbstbild der Vp erhält, als vielmehr ein Abbild der allgemeinen kulturellen Norm : sozial negativ bewertete Variablen werden auch in der Selbstbeurteilung negativ eingeschätzt, sozial akzeptable 'Eigenschaften' werden demgegenüber vorrangig als 'charakteristisch' eingestuft. Die 'social-desirability' (ferner als s-d abgekürzt) ist also eine Variable, die den items selbst zukommt und welche - wie derartige Untersuchungen aufzuzeigen scheinen - die Reaktion auf diese items wesentlich bestimmt.

Der Grad an s-d, der jedem item zukommt, wird in einem rating-Verfahren festgestellt : eine Urteiler-Gruppe schätzt jedes item hinsichtlich seiner s-d ein. Handelt es sich dabei um items eines Q_s-sort-Verfahrens, wird die s-d auf einer Skala eingeschätzt, die derjenigen des Q_s-sort's selbst entspricht, so daß man nun vergleichen kann, wie dieselben

---

1) Edwards 1953, 1955; Edwards et.al. 1959; Corah et.al.1959; Kelleher 1958; Buss 1959; Cowen,Tongas 1959; Goodstein, Heilbrunn 1959; Rozynko 1959; Boe-Kogang 1963; dazu kritisch : Bills 1959; Taylor 1959.

items einmal hinsichtlich ihrer s-d beurteilt und in welcher Weise sie zum anderen für eine Selbst-Charakterisierung benutzt werden. Dieser Vergleich aber wird folgendermaßen vorgenommen : der Mittelwert aller 'Urteile' über ein item, als Index seiner s-d, wird korreliert mit jenem durchschnittlichen Stellenwert, der demselben item in den Selbsteinschätzungen einer anderen Gruppe zukommt. Es ergeben sich bei den verschiedenen Untersuchungen Korrelations-Koeffizienten im Werte von o,5 bis o,95 [1].

Diese relativ hohen und signifikanten Korrelationen werden dahingehend interpretiert, daß tatsächlich die Art der Einschätzung der items zum Zwecke der Selbstbeschreibung derjenigen Einstufung entspricht, welche den Grad der s-d dieser items bemißt. Die items werden daher für die Aufgabe eines self-reports als ungeeignet angesehen, und es wird gefolgert, daß man innerhalb eines Verfahrens nur items mit demselben s-d-Wert verwenden dürfe, will man nicht ein Abbild der den items zukommenden s-d-Werte, sondern eines der Person erhalten.

Dieses Vorgehen der Korrelation von Durchschnittswerten sowohl auf Seiten des Q_s-sorts wie der s-d-Einschätzungen muß jedoch in mehrfacher Hinsicht kritisiert werden.

1. Aufgrund des solipsistischen Charakters der Daten eines Q_s-sort's ist es methodologisch nur mit Einschränkung vertretbar, Mittelwerte der scores pro item (oder Eigenschaft) über die Individuen hinweg (also inter-individuell) zu bilden. Es braucht hier nur auf die eingehende Diskussion des relationalen Charakters dieser Daten verwiesen zu werden.

2. Entscheidender ist, daß eine solche Mittelwertsbildung für diese Fragestellung prinzipiell irrelevant ist. Erst die Variabilität der Urteile bzw. Selbsteinschätzungen hinsichtlich eines items kann den Einfluß der Variable s-d adäquat widerspiegeln : die Mittelwertsbildung erweckt näm-

---

1) siehe Fußnote 1) der Vorseite.

lich demgegenüber den völlig falschen Eindruck, als hätten a l l e Vpn ihre Einschätzung nach der s-d ausgerichtet. Doch kommt es allein darauf an, ob die items dazu geführt haben, daß man sie d u r c h g ä n g i g nach ihrer s-d einschätzt : je größer aber die Variabilität der Reaktionen auf die items im Q_s-sort, um so weniger wirksam ist der Einfluß von s-d. Der Informationsverlust, der mit der Mittelwertsbildung verbunden ist, läßt es also nicht zu, derartige Durchschnittswerte als Mittel der Kritik eines Verfahrens zu benutzen. Vielmehr ist umgekehrt der Grad an Variabilität (Streuung) der Urteile ein Kriterium gerade dafür, daß das Verfahren inter-individuell hinreichend differenziert. Die Variabilität wird somit zu einem Kriterium für die Gültigkeit des Verfahrens.

3. Das gleiche gilt für die Mittelwertsbildung, welche zu dem Index von s-d pro item führt : auch hier wird fälschlich der Eindruck erweckt, als sei der Grad der s-d für alle Vpn (also für alle 'Urteiler') in eindeutiger Weise der gleiche. Es könnte aber die Variabilität auch dieser Urteile so groß sein, die Mehrdeutigkeit eines items hinsichtlich dessen, was man für 'sozial akzeptabel' hält oder nicht, so ausgeprägt sein, daß ein einheitlicher Mittelwert sowohl unberechtigterweise einen Gruppenkonsensus vortäuschte, wie damit zugleich verschleiern würde, daß eine allgemeine Norm hinsichtlich dieses bestimmten items gar nicht existiert.

4. Die Korrelation aber nun dieser Mittelwerte läßt, wenn man sie im obigen Sinne versteht, allenfalls folgende Interpretation zu : Gedachte Personen, welche im Vergleich zu einer bestimmten Bezugsgruppe in ihrer Selbstbeurteilung die items auf eine jeweils 'durchschnittliche' Weise einschätzen, entsprechen in der so entstandenen Struktur ihres Eigenschaftsgefüges den Vorstellungen solcher - ebenfalls z u d e n k e n d e r - Leute, welche 'durchschnittliche' Ansichten haben über den Grad an social-desirability bei eben diesen items. Es bedeutet dies also, daß solche gedachten Personen, die sich selbst als durchschnittlich hinsichtlich ihrer Eigenschaften begreifen, dem entsprechen, was andere gedachten Personen

(hier die 'Urteiler' für s-d) als durchschnittlich ansehen.

Es zeigt sich, daß diese - allein zulässige - Interpretation derartiger Mittelwertskorrelationen für das Problem, das mit dem Begriff der s-d gestellt ist, völlig unergiebig ist. Zugleich aber ergibt sich die weit wichtigere Frage, ob das Problem selbst richtig gestellt ist. Es gilt zunächst einmal näher zu analysieren, ob es überhaupt eine befremdende Tatsache ist, wenn Vpn sich in ihrem Selbstbild (für den Fall, d a ß sie es tun, - was jedoch mit einer derartigen Korrelation überhaupt nicht auszumachen ist) den allgemeinen kulturellen Normen relativ entsprechend einschätzen. Das Faktum, daß die real-self und ideal-self-scores bei den einzelnen items teilweise zusammenfallen mit den s-d-Werten, welche hier als 'allgemeine kulturelle Norm' aufgefaßt seien, läßt sich nämlich auch innerhalb folgenden Zusammenhangs begreifen.

1. Das Ich-Ideal ist bestimmt durch die kulturelle Norm, die ihrerseits die Grundlage bildet für die Einschätzungen der 'social-desirability'.

2. Sogenannte 'gut angepaßte' bzw. 'emotional nicht gestörte' Personen entsprechen nicht nur weitgehend ihrem Ich-Ideal, sondern eben deshalb auch der kulturellen Norm und damit zugleich per definitionem der 'social desirability'.

3. Die sozialen Stereotypen, die der Einschätzung von s-d als zugrundeliegend gedacht werden müssen, und die kulturellen Normen, welche über den Weg der Über-Ich-Orientierung entscheidend verhaltenssteuernd wirken, dürfen also nicht als unabhängig voneinander betrachtet werden. Vielmehr unterliegen hier soziales Verhalten, soziales Wahrnehmen und Urteilen den gleichen Bedingungen einer bestimmten kulturellen Umwelt. Die kognitiven Inhalte solcher soziokultureller Normen stehen sowohl in einem bedingenden, wie bedingten Verhältnis zum Verhalten. Im Verhalten selbst aber - im Gegensatz zur Norm - ist mit einem weiten Spielraum inter-individueller Unterschiede zu rechnen.

Anstelle der Behauptung von 'social-desirability', welche

als den items 'anhaftend' gedacht wird, ergibt sich somit die umgekehrte Frage : ist ein item so allgemein gefaßt, daß es im Prinzip auf jedes Glied einer solchen kulturellen Umwelt in einem relativ einheitlichen Grade zutrifft; oder aber ist es spezifisch genug gefaßt, daß es eine inter-individuelle Differenzierung hinreichend ermöglicht. Sind derartige items, anhand welcher inter-individuelle Unterschiede zum Ausdruck gelangen können, gegeben, so bleibt nicht zu fragen, welcher Grad an s-d ihnen generell zugeordnet werden kann, sondern ob diese erwarteten inter-individuellen Unterschiede in den Reaktionen tatsächlich eintreten. Diese Frage aber beinhaltet nicht die Annahme einer 'Eigenschaft', welche den items zukäme, sondern verweist auf den Umstand, daß Vpn entweder ihrer Selbsteinsicht entsprechend oder aber in Orientierung an der sozialen Norm die items beantworten. 'Social desirability' erweist sich so als ein Phänomen, das nicht gewissermaßen eine 'Disposition' von items bezeichnet, sondern vielmehr Verhaltensformen von Vpn charakterisiert. Das aber besagt, daß der Begriff irreführend und das Problem falsch gestellt ist.

Ein item, das nicht zu inter-individuell verschiedenartigen Reaktionen führt, ist diagnostisch unergiebig [1]. Eine genügend große inter-individuelle Streuung aber verweist darauf, daß die sog. 'social-desirability' keine Rolle spielt, mag es dabei auch einzelne Vpn geben, die sich stärker an der Norm als an ihrem Verhalten bei der Selbsteinschätzung ausrichten.

Aufgrund dieser Überlegungen nun lassen sich - in erklärter Abgrenzung gegen das Konzept der 'social desirability' - Kriterien herleiten, anhand deren sich Teilaspekte der Gültigkeit eines Q_s-sort-Verfahrens untersuchen lassen.

---

1) Auch dabei muß es sich nicht um einen bestimmten Grad an s-d handeln, sondern das in dem item bezeichnete Verhalten kann so unspezifisch sein, daß es auf alle Vpn in relativ gleicher Weise zutrifft.

8. Hypothesen zur Validität und Stabilität des hier verwendeten Q_s-sort-Verfahrens

Als ein besonders wichtiges Kriterium für die Brauchbarkeit eines Q_s-sort's ist oben abgeleitet worden, daß das Verfahren - statt gleichartiger sozialer Stereotype - eine hinreichende inter-individuelle Streuung in den Reaktionen hervorbringen müsse. Denn im Verhalten drückt sich der individuelle Sozialisierungsprozeß aus, der zwar an den sozio-kulturellen Normen orientiert ist, jedoch diese nicht notwendig reproduziert. Die Bezugsnormen für das Verhalten drücken sich vielmehr - per definitionem - im Ich-Ideal aus, so daß hier eine hohe inter-individuelle Übereinstimmung nicht zu Lasten des Verfahrens gerechnet werden kann, sondern vielmehr ein Kriterium für die experimentelle Repräsentanz der theoretischen Realität darstellt, also erwartet werden muß.

Daraus folgt, daß insbesondere bei sgn. 'gut-angepaßten' Vpn das real-self wenigstens teilweise dem ideal-self entspricht, was nur den A n s c h e i n erweckt, als habe social desirability hier als Störfaktor eingewirkt.

Das Wunsch-Ich wird, da es per definitionem auch jene Verhaltens-Wünsche umfaßt, welche der sozialen Anpassung an bestehende Normen entgegenstehen, am stärksten von der kulturellen Norm abweichen. Es wird inter-individuell stärker variieren als das ideal-self, weil es nicht wie dieses durch allgemeine Normen vereinheitlicht wird. Doch werden diese Wünsche nicht eine solche inter-individuelle Verianz aufweisen, wie das Verhalten selbst : da diese 'Ebene' - teilweise ein Gegenbild zur sozialen Norm - gerade auch solche Wünsche enthält, welche durch die soziale Norm allgemein sanktioniert werden und darin sich in inter-individuell ähnlicher Weise gerade an dieser Norm bemessen lassen.

Von diesen theoretischen Voraussetzungen ausgehend, läßt sich folgende Hypothese herleiten :

I. Die Einschätzungen aller Vpn zu einem bestimmten item, resp. zu einer bestimmten Variablen (hier werden die 8 zugehörigen items in ihrem Median zusammengefaßt, vgl.S.204) werden im real-self am stärksten voneinander abweichen, also die größte i n t e r -individuelle Streuung aufweisen; die i n t e r -individuelle Varianz der Einschätzungen wird im ideal-self am geringsten, im Wunsch-Ich von mittlerer Größe sein.

Wenn hier - gegen die Untersuchungen zur sog. 'social desirability' - das Hauptgewicht auf die Varianz der Urteile gelegt wurde, so sollte es nicht unterlassen werden, die Methodik dieser kritisierten Untersuchungen auch auf das hier vorliegende Material anzuwenden. Wenn nämlich behauptet wird, daß die (irreführenden) Ergebnisse solcher Untersuchungen allein auf die inadäquate Methode zurückzuführen sind, so müßten sich mit eben derselben Methode auch hier Ergebnisse einstellen, die scheinbar den Nachweis der 'störenden' Wirkung einer nicht kontrollierten 'social desirability' erbringen. Die besagte Methodik besteht darin, die Varianz der Urteile unberücksichtigt zu lassen und allein die Mittelwerte der Urteiler-Gruppe (social-desirability) und der Versuchsgruppe (real-self) aufeinander zu beziehen; dies mit dem Ergebnis, daß für die einzelnen Variablen die Mittelwerte der Selbsteinschätzungen zusammenfallen mit den Mittelwerten der Urteile hinsichtlich s-d. Bei einem analogen Vorgehen ergibt sich für unser Material, in welchem das ideal-self die 'Norm' repräsentiert, folgende Hypothese :

II. Der Mittelwert aller Einschätzungen zu einer Variablen im real-self entspricht dem Mittelwert aller Einschätzungen derselben Variablen im ideal-self.

Wenn sich beide Hypothesen bestätigen lassen, so erfährt die Argumentation des vorliegenden Abschnitts eine empirische Bestätigung : Übereinstimmung der Mittelwerte im obigen Sinne (Hypothese II) besagte nichts über das Wirken von 'social desirability'; denn anhand der Varianz der Urteile (Hypothese I) wäre nachgewiesen worden, daß die Reaktionen im

Bezugssystem 'real-self' nicht den Normen entsprechen, die im 'ideal-self' zum Ausdruck gebracht werden.

Noch ein weiterer Gesichtspunkt kann herangezogen werden, welcher ebenfalls innerhalb des Datenmaterials des Q_s-sort's ein Kriterium für die construct-Validität des Verfahrens erschließt. Dieser Gesichtspunkt ergibt sich, wie im obigen Fall (Hypothese I), aus den Definitionen der 3 Urteils-Ebenen. Ein weiterer wesentlicher Unterschied zwischen real-self einerseits und ideal-self und Wunsch-Ich andererseits nämlich läßt sich dahingehend charakterisieren, daß die Einschätzung des eigenen Verhaltens sehr unterschiedliche situative Determinanten mit berücksichtigen muß [1], da das Verhalten je nach Situation sehr stark variieren kann. Demgegenüber kann in der Einschätzung von ideal-self und Wunsch-Ich von situativen Determinanten gerade abgesehen werden, da es hier nicht um die Einschätzung konkret-verhaltensmäßiger Anpassung geht, sondern konsistente Beurteilungsprinzipien leitend werden. Mein Wunsch nach Aggressivität kann ungeachtet der Situationen durchgehend stark sein; im Ich-Ideal mag prinzipiell jegliche Aggression abgelehnt werden, während im Verhalten - je nach Situation - sehr unterschiedliche Formen der Aggressivität bei einer Person beobachtet werden können. Diese Unterschiede zwischen den Bezugs-'Ebenen' müssen demnach darin zum Ausdruck kommen, wie die Vp auf den einzelnen Ebenen die je 8 zu einer Variablen gehörigen items einschätzt, so daß sich folgende Hypothese ergibt :

III. Die i n t r a -individuelle Varianz der je 8 Urteile zu einer Variablen wird im real-self am größten und bei derselben Person im ideal-self und Wunsch-Ich entsprechend geringer sein.

Läßt sich aber eine sign. Differenz zwischen der intra-individuellen Streuung im real-self einerseits und im ideal-self und Wunsch-Ich andererseits nachweisen, so ist damit

1) Die 8 zu einer Variablen zugehörigen items sind - wie bereits dargelegt - so konzipiert worden, daß sie verschiedene Situationen umfassen.

die obige Hypothese (III) noch nicht vollständig bestätigt, da einem solchen Ergebnis auch ein anderer Bedingungszusammenhang zugrundeliegen kann. Die in dieser Hypothese ausschlaggebende Eigenart der intra-individuellen Streuung im real-self nämlich kann - sollte sie sich ergeben - u.U. allein auf einen Positionseffekt zurückgeführt werden : das real-self wird von jeder Vp als erstes eingeschätzt, die items sind hier - im Gegensatz zu den beiden weiteren Urteils-Bezugssystemen - noch nicht bekannt. Bei ideal-self und Wunsch-Ich könnte sodann allein die nunmehr erworbene Geübtheit mit dem Verfahren dazu führen, daß die items sehr viel konsistenter, nämlich stärker ihrer logischen Zugeordnetheit nach, eingeschätzt würden [1]. Sollte sich ein solcher Positionseffekt nachweisen lassen, so entbehrt ein nicht unwichtiger Aspekt der construct-Validität des Verfahrens seiner empirischen Grundlage.

Es ist daher im Abstand von einer Woche die Einschätzung des real-self noch einmal wiederholt worden ('real(2)', s.S. 191), so daß nun geprüft werden kann, ob die obige Hypothese (III) auch für das real(2) zutrifft oder ob die Vertrautheit mit den items dazu führt, daß die intra-individuelle Streuung bei der Einschätzung des eigenen Verhaltens hier derjenigen von Ich-Ideal und Wunsch-Ich entspricht. Soll also die Annahme eines Positionseffektes zurückgewiesen werden können, muß sich folgende Hypothese bestätigen lassen :

IV. Die intra-individuelle Varianz der je 8 Urteile zu einer Variablen ist, unabhängig von der Position, sowohl für real (1), wie für real (2) größer als für das ideal-self und Wunsch-Ich. Zwischen real (1) und real (2) besteht kein Unterschied hinsichtlich der intra-individuellen Streuung.

Anhand dieser Daten des real-self(2) kann zugleich festgestellt werden, wieweit das Verfahren gleichartige Daten hervorbringt, wenn es (analog zur Re-Test-Methode) im Abstand einer Woche wiederholt wird. Die Korrelation zwischen den Meßwerten der ersten

1) Es sei hier noch einmal vergegenwärtigt, daß die Vp 88 items in einer zufälligen Reihenfolge (nacheinander) einzuschätzen hat, die sich in 11 Untergruppen von je 8 items (bei 11 Variablen) untergliedern lassen; bei der ersten Einschätzung (real-self) ist also noch nicht bekannt, welche items überhaupt zu beurteilen sind.

und einer späteren Durchführung eines Tests bei denselben Vpn ist als die 'Stabilität' des Tests definiert. Auch hier kann die Korrelation zwischen real(1) und real(2) bestimmt werden. Da wir jedoch nicht, wie es der üblichen Bestimmung der Stabilität entspräche, zwei Meßreihen bei einer Gruppe von Vpn korrelieren, sondern für jede einzelne Vp einen Korr-Koeff. (real(1)/real(2)) erhalten, haben wir hier die V e r t e i l u n g dieser 34 Korr-Koeff. zu analysieren. Wir erwarten, daß diese Korr-Koeff. zwischen real(1) und real(2), die wir zur Abgrenzung gegen die übrigen Korrelationen zwischen Q_s-sort-Daten 'Stabilitätskoeffizienten' nennen wollen, hoch sind, daß sie insbesondere höher sind als die Korr-Koeff. zwischen real und ideal oder real und Wunsch etc. Ferner erwarten wir, daß diese Stabilitätskoeffizienten durchgehend hoch sind, während wir für den Vergleich z.B. zwischen real und ideal erhebliche inter-individuelle Differenzen (nämlich Unterschiede hinsichtlich des 'Ebenen-Gefüges') erwarten. Somit sind folgende Hypothesen aufzustellen :

V. Die Korr-Koeff. zwischen real(1) und real(2) sind höher als diejenigen z.B. zwischen ideal und Wunsch.

VI. Die Streuung der Stabilitätskoeffizienten ist geringer als die Streuung, welche die Verteilung z.B. der Korr-Koeff. zwischen real und ideal aufweist.

Anhand solcher Stabilitätskoeffizienten läßt sich noch ein weiterer Gesichtspunkt überprüfen. Die obige Erwartung einer sign. geringeren inter-individuellen Streuung im ideal-self und Wunsch-Ich war durch die Annahme begründet worden, daß in diesen Bezugssystemen anhand eines einheitlichen Beurteilungs- p r i n z i p s geurteilt würde, daß hier größere kognitive Klarheit darüber bestünde, was man 'sollte' oder nicht bzw. was man 'möchte' oder nicht; während die Beurteilung des eigenen Verhaltens nach den verschiedenartigen Verhaltenssituationen zu differenzieren sei. Diese größere intra-individuelle Streuung in der Einschätzung des real-self könnte jedoch stattdessen darin begründet sein, daß für die Beurteilung der Vielfalt von Verhaltensäußerungen eben diese kognitive Klarheit nicht besteht. Eine solche Urteilsunsicherheit aber würde die Zuverlässigkeit des Verfahrens schmälern und sich in einer geringen Stabilität äußern. Es muß daher aufgewiesen werden können, daß

es nicht eine Urteilsunsicherheit ist, welche zu der größeren intra-individuellen Streuung im real-self führt. Dieses ist mit Hilfe der Stabilitätskoeffizienten möglich : Urteilsunsicherheit nämlich müßte gepaart sein mit geringer zeitlicher Stabilität der Urteile; je größer die intra-individuelle Varianz im real-self, desto geringer müßte der Stabilitätskoeffizient sein. Soll stattdessen nachgewiesen werden, daß die hohe intra-individuelle Streuung im real-self nicht auf einen Mangel an kognitiver Klarheit zurückgeführt werden kann, so daß die Hypothese an Wahrscheinlichkeit gewinnt, in dieser Varianz käme die situative Bedingtheit des Verhaltens zum Ausdruck, so muß folgende Hypothese bestätigt werden :

VII. Es besteht kein korrelativer Zusammenhang zwischen dem Grad der intra-individuellen Varianz im real(1) und der Größe der Stabilitätskoeffizienten.

Sollten sich diese verschiedenen Hypothesen bestätigen lassen, so wäre zwar nicht empirisch gesichert worden, was das hier verwendete Q_s-sort-Verfahren 'tatsächlich' mißt, doch wäre damit in verschiedenen Punkten eine befriedigende 'Repräsentanz' der 'theoretischen' durch die 'experimentellen' Sätze nachgewiesen worden. [1)] Zumindest für einzelne konstruktive Elemente des theoretischen Ansatzes hätte sich das Verfahren als angemessenes Korrelat bewährt.

Die inhaltliche Repräsentanz psychischer Realitäten wie 'erlebte Aggressivität' oder 'Aggressions-Wünsche' konnte nicht im einzelnen nachgeprüft werden (vgl.S.167f). Sie wird - daran sei abschließend noch einmal erinnert - geprüft durch die Interrelationen zwischen Q_s-sort- und WP-Daten, - was notwendig die Einschränkung mit einschließt, daß die Ergebnisse nicht über die besondere Versuchsanordnung hinaus generalisiert werden können. Diese Versuchsanordnung und die experimentellen Ergebnisse sind im folgenden darzustellen.

---

1) vgl. die Einführung dieser Begriffe auf S. 99 ff

D R I T T E R T E I L

D a s E x p e r i m e n t

VII. Die Durchführung des Experiments

1. Die Versuchsanordnung

Da im Zweiten Teil und insbesondere in Kap. V, Abschnitt 2 alle Gesichtspunkte hergeleitet wurden, die zur Planung des Experiments im einzelnen geführt haben und dieses zugleich theoretisch begründen, kann nunmehr auf eine eingehendere Erläuterung verzichtet und das Experiment in seinem Verlauf skizziert werden.

Die Untersuchung wurde im ersten Halbjahr 1966 durchgeführt. Als Versuchspersonen waren alle weiblichen Neuimmatrikulierten der Philosophischen Fakultät im WS 1965/66 und teilweise auch vom SS 1966 angeschrieben worden : aus rund 100 Adressaten rekrutierte sich schließlich eine Versuchsgruppe von 34 Vpn. Diese stellt insofern nicht eine 'Zufallsstichprobe' dar, als nur jene Studenten herangezogen wurden, die positiv geantwortet hatten. Es ist aber bereits dargelegt worden, weshalb nicht die Frage der Repräsentativität, sondern nur die der relativen Homogenität der Versuchsgruppe hier eine Rolle spielt (vgl. S.124).

Es wurden nur weibliche Versuchspersonen herangezogen, weil der TAT für beide Geschlechter einen verschiedenen Satz von Bildern enthält, so daß - hätten wir weibliche u n d männliche Vpn genommen - nicht konstante Bedingungen gewährleistet wären. Eine weitere Rolle spielte die Annahme, daß weibliche Vpn der hier gebotenen Aufgabenstellung, aber auch einem weiblichen Versuchsleiter gegenüber offener sein würden als ihre männlichen Kollegen. Die Philosophische Fakultät ist von der Überlegung ausgehend gewählt worden, daß die Studenten hier sprachlichem Material und auch sprachlicher Äußerung (Geschichten) gegenüber aufgeschlossener sein, wie aber auch der Aufforderung zu fantasiehaften Vorstellungen leichter und mit weniger Vorbehalt entsprechen könnten als beispielsweise Mediziner oder Naturwissenschaftler.

Da in diesem Experiment die Erhebung von self-reports eine entscheidende Rolle spielt, mußten wir durch eine entsprechende Vorinstruktion versuchen, eine hinreichende Motivation für Offenheit und 'Ehrlichkeit' zu erzielen. Es wurde an das 'wissenschaftliche Ethos' appelliert, indem dargelegt wurde, daß allein eine sachliche Einstellung gegenüber dem Vorhaben und eine entsprechende Offenheit in der Reaktion überhaupt zu einer wissenschaftlichen Ergiebigkeit des Experiments führen könne [1]. Anonymität war dabei zugesichert. - Der Vorinstruktion ging ein allgemeines einleitendes Gespräch darüber voraus, welche Erfahrungen mit bzw. welche Kenntnisse über die Psychologie die Vp bereits gewonnen habe.

Die Untersuchung verteilte sich auf 2 Tage mit genau einer Woche Abstand. Jede einzelne Sitzung dauerte mindestens 3, teilweise 4 Stunden.

Die Wunschprobe wurde in der von Wilde vorgeschlagenen Standardform gegeben [2] : die 66 items waren auf Kärtchen von 1 x 6 cm maschinenschriftlich aufgedruckt. Die 88 items des Q_s-sort-Verfahrens lagen ebenfalls in Maschinenschrift auf einzelnen Karten (3 x 9 cm) vor.

Da die Wunschprobe als Gegenstand der Untersuchung das wichtigste der angewandten Verfahren darstellt, wurde sie als erste dargeboten. Die Verteilung des Untersuchungsprogramms auf 2 Sitzungen war folgendermaßen :

1.Tag:	Wunschprobe	2.Tag:	Erinnerungsexperiment
	Q-sort: real-self(1)		Q-sort: real-self(2)
	ideal-self		T A T
	Wunsch-Ich		

Zur Durchführung der Versuche läßt sich allgemein sagen, daß die Vpn durchgehend offener und der Aufgabenstellung gegenüber aufgeschlossener waren, als wir es erwartet hatten.

1) Alle Instruktionen finden sich im genauen Wortlaut im ANHANG.
2) Die Instruktionen sind allerdings abgewandelt worden: die an Märchen anklingende Vorstellung "Sie seien gestorben" sollte Erwachsenen nicht zugemutet werden (siehe ANHANG).

Für die Wunschprobe läßt sich - mit wenigen Ausnahmen - der Wilde'sche Satz zitieren : "Gewöhnlich beginnt - nach einem kleinen Moment des Staunens - oder des kritischen Abwägens - der Proband mit einer gewissen Freude das Spiel." (1950, S.214) [1] Nur in wenigen Fällen gab es Schwierigkeiten in der Verbalisierung der Begründungen. - Insbesondere das Q_s-sort-Verfahren wurde zumeist sehr gewissenhaft durchgeführt, die Entscheidungen immer wieder überprüft; auch nach erledigter Aufgabe(vorgeschriebene Verteilung) wurde zumeist noch einiges korrigiert, jedenfalls das endgültige Gefüge der items noch einmal eingehend überprüft. Dieses gilt insbesondere für die allererste Durchführung (real (1)) : bei den weiteren Bezugssystemen wirkte es sich auf das Tempo und auch in einigen Fällen auf den Grad an kritischer Sorgfalt aus, daß Methode wie items nunmehr bekannt waren.

Von wenigen Ausnahmen eines Desinteresses und einer Ausnahme ausdrücklicher Ablehnung abgesehen, wollten alle Vpn über ihre persönlichen Ergebnisse, wie die allgemeinen Ergebnisse der Untersuchung informiert werden. Von dem Angebot einer 'Aufwandsentschädigung' in Höhe von 10 DM wurde in 2 Fällen in voller Höhe Gebrauch gemacht.

2. Das Datenmaterial und seine statistische Verarbeitung

a) Die Auswertung der Wunschprobe

Da es für die Wunschprobe bisher keinen irgendwie erprobten oder schon bewährten Auswertungsmodus gibt, mußte eine Methode speziell für unsere Zwecke entwickelt werden. Sie wurde allein an den Erfordernissen des Experiments und seiner Fragestellung ausgerichtet. Das bedeutet vor allem, daß es sich nicht um eine Auswertung im Sinne diagnostischer 'Interpretation' handeln kann : es wird also nicht etwa vom 'manifest-

---

1) Es drückt sich dies z.B. auch darin aus, daß eine Reihe von Vpn diese Methode mit ihren Bekannten ausprobierten (dies allerdings der Instruktion entgegen, sich in der Zeit zwischen den beiden Sitzungen nicht weiter mit dem Experiment zu beschäftigen).

funktionalen Zeichengehalt' auf dessen 'latent-funktionale Bedingungen' (Heiß, 1951a) geschlossen. Es ist ja vielmehr gerade der Ansatzpunkt dieser Arbeit, ein solches 'Interpretieren' in Frage zu stellen und eben nicht bereits irgend eine Art des Funktionierens bei einem projektiven Verfahren vorauszusetzen.

Wie aber kann man ein solches Verfahren in 'scores' übertragen, die einerseits eine Interpretation des Testprotokolls nicht voraussetzen und zugleich eine Subjektivität des Urteilens auf der Seite des Auswerters ausschließen, nämlich zu einer Art 'objektiver' Quantifikation führen?

Eine Testinterpretation im diagnostischen Sinne wird hier dadurch ausgeschlossen, daß allein der sprachlich manifeste Inhalt der Begründungen in der WP zur Grundlage gemacht wird : dieser nämlich wird formalisiert, d.h. in ein anderes 'Sprachsystem' übertragen. Dieses andere Formulierungssystem gründet seinerseits auf einer formal-inhaltlichen Auszählungs-Methode, wie sie in der Forschung schon seit längerem, insbesondere für den TAT, angewendet wird. Alle Untersuchungen, welche den Ausdruck von Motiven in projektivem Material zum Gegenstand haben, müssen mit irgendeiner Auszählungsmethode arbeiten; entsprechend haben z.B. alle in Kap. II behandelten Experimente diese Grundlage, wenn sie auch nirgendwo näher ausgeführt wird [1] (mit Ausnahme allerdings von McClelland et.al. 1953).

Eine andere Vorgehensweise noch bestünde in einer weiteren formal-inhaltlichen oder einer rein formalen Auswertungsmethode, also einer solchen, die von allem Inhaltlichen völlig absieht :

---

1) Es gehen alle diese Formen einer Quantifizierung letztlich auf das ursprüngliche Murray'sche Auswertungssystem zurück (Murray 1943) und sind vergleichbar auch mit dem noch stärker gegliederten Verfahren von Aron (1949); nur greifen diese Untersuchungen im Rahmen einzelner wissenschaftlicher Fragestellungen jeweils bestimmte isolierte Aspekte heraus. Versuche einer näheren Ausarbeitung solchen Vorgehens finden sich bei Fine (1955); Stone (1956); Pine (1960); Hafner, Kaplan (1960); Gottschalk, Gleser, Springer (1963).

im ersteren Fall würde man in der WP etwa den Quotienten der positiv getönten Adjektive in den Begründungen zu den negativ getönten Adjektiven und im letzteren Fall einen Verb-Adjektiv-Quotienten (vgl. Balken u. Massermann 1940) bilden können. Es ist jedoch dieses Vorgehen bisher theoretisch am wenigsten fundiert und stellt somit kaum empirisch bereits begründete oder naheliegende Hypothesen bereit. Es kann aber heute ebensowenig bereits als erschöpft angesehen werden, und wird wahrscheinlich auch für die WP fruchtbar gemacht werden können, insbesondere von einem weniger rein sprach-logischen, als vielmehr 'person-centered' [1] Ansatz aus (vgl. dazu auch Holt 1958). Allerdings konnte ein solcher Versuch im Rahmen dieser Arbeit nicht mehr zusätzlich realisiert werden und ist auch deshalb kaum angezeigt, da es hier allein um den i n h a l t l i c h e n Ausdruck bestimmter M o t i v e geht.

Die Methode einer formal-inhaltlichen Umformulierung war für die Zwecke des hier vorgenommenen Experiments (mit seinem Katalog bestimmter, der Untersuchung vorangestellter Variablen, s.S.146) eigens zu entwickeln. Für die WP sind die B e g r ü n d u n g e n zu der jeweiligen Wahl der 66 items Grundlage einer solchen inhaltlichen Transformation. Erschwerend ist hier, daß diese Begründungen mehr oder weniger ausführlich und teils nur in Stichworten gegeben werden. Sie sind oft so sehr nur Abkürzungen [2] dessen, was die Vp im Augenblick der Reaktion weit ausführlicher 'dachte', daß nicht allein der Schluß auf das indirekt in der Antwort Mitgemeinte nötig wird, sondern teilweise auch auf die in der Begründung implizierte psychische Dimension [3] (also auf den 'l a t e n t funktionalen Zeichengehalt', vgl.S.192f). Die

1) Solche stärker von psychologischen als sprachlich-formalen Aspekten ausgehende Ansätze finden sich etwa bei Schafer (1948), Rapaport (1946) und in letzter Zeit bei Wyatt (1958), Kagan (1961) und Shneidman (1961).
2) 'Abkürzung' in einem formalen Sinne und nicht in dem hier ebenfalls relevanten Sinne der entstellenden Äußerung des wirklich Gedachten.
3) vergleiche das Beispiel im nächsten Absatz.

Grenze zwischen einer sich logisch ergebenden Erweiterung des sprachlichen Inhalts einer stichwortartigen Antwort einerseits und einer 'Interpretation' in Richtung auf den latenten psychischen Gehalt andererseits ist dabei nur sehr schwer einzuhalten.

Es sei dies an einem Beispiel verdeutlicht, etwa der Begründung der positiven Wahl des item's 'Rose' : "sie hat Dornen". Erweitert man als Auswerter diese Antwort zu der inhaltlich ergiebigeren Form "sie hat Dornen und kann damit stechen", so ist dabei schwer zu entscheiden, ob es sich um eine bloße sprachliche Extrapolation handelt (von der Annahme ausgehend, was denn anders bemerkenswert an 'Dornen' sei, als daß sie stechen; also von einer ganz bestimmten Konnotation ausgehend); oder ob es sich um eine psychologisch interpretative (und zudem 'intuitive') Schlußfolgerung handelt derart, daß man bei dem Stichwort 'Dornen' sogleich Aggressivität vermutet und diese nicht formal-sprachliche, sondern 'psychologische' Auffassung durch den vorgeblichen formal-sprachlichen Kontext dieser Kurzantwort im nachhinein rechtfertigt. Diese Grenze zwischen einer formalen und einer inhaltlichen Erweiterung einer solchen Antwort ist teilweise schwer zu ziehen. Entscheidend ist, daß ein scoring-system entwickelt wird, ohne daß Hypothesen über das Verhältnis von Index und Indiziertem bei der WP leitend sind. Es ist unbedingt zu vermeiden, daß derartige Hypothesen beim Signieren der Testantworten ausschlaggebend werden. Die Testantworten sollen 'transformiert', nicht 'interpretiert' werden.

In ein solches Auswertungsschema aber gehen notwendig gewisse Präferenzen, also das 'bias' des Autors, mit ein : erst im nachhinein, nämlich in der Anwendung, kann sich zeigen, wieweit es tragfähig ist und wieweit nicht. Doch müssen diese subjektiven Determinanten ihrerseits bestimmbar und das heißt nachprüfbar sein; die Auswertungskriterien müssen objektivierbar, also auch in ihrer Benutzung wiederholbar sein.

Zur Orientierung ist hier, ähnlich dem Vorgehen der deutschen Herausgeber des PFT (Hörmann, Moog 1957), ein (der Intention

nach) repräsentativer Katalog von Testantworten und ihrer Signierung zusammengestellt worden (siehe ANHANG, S. 316) Dieser Katalog ist allein ausgerichtet auf die Zwecke dieser Untersuchung und basiert ausschließlich auf dem Antworten-Material der Versuchsgruppe [1]. Seine Aufgabe ist es, den sonst üblichen, aber sehr viel schwieriger zu realisierenden Umweg über den 'Urteiler-Konsensus' als Mittel zur Objektivierung der Auswertung zu ersetzen : auch ein solcher setzte nämlich ein einheitliches Beurteilungskonzept bereits voraus und prüfte zwar dessen Anwendbarkeit, doch nicht dieses selbst [2]. Stattdessen ist der Versuch gemacht worden, das Konzept so zu präzisieren, daß die Annahme seiner praktischen Anwendbarkeit gerechtfertigt erscheint. Die 'Bewährung' des Auswertungsschemas jedoch läßt sich hier nur daran bestätigen, ob die Daten, die es hervorbringt, im Rahmen der theoretischen Hypothesen zu den erwarteten Ergebnissen führen. [3]- Es sei noch einmal darauf verwiesen, daß dies Auswertungsschema, sowie formale Gesichtspunkte seiner praktischen Anwendung, im ANHANG nachzuschlagen sind.

Wir wollen nun das Datenmaterial selbst, das aus diesem Auswertungsschema hervorgeht, näher charakterisieren (s.Tabelle im ANHANG). Jede Vp erhält für jede der 11 Variablen (vgl. S.146) 2 scores : nämlich einen positiven und einen negativen

1) Da dieser Katalog an den I n h a l t e n der Testreaktionen ausgerichtet ist, so wie sie bei dieser Gruppe auftreten, müßte er für andere Versuchsgruppen wahrscheinlich modifiziert werden.
2) Ein Urteiler-Konsensus würde allein darauf deuten, daß das Auswertungsschema verständlich und 'praktikabel' ist. Ob es den (noch festzustellenden) A u s s a g e -Möglichkeiten des Tests angemessen ist, würde durch die in dem Konsensus angezeigte 'Allgemeinverständlichkeit' des Schemas nicht berührt. Darauf, diese 'Allgemeinverständlichkeit' des vorgeschlagenen Auswertungsschemas eigens zu prüfen, wurde hier verzichtet.
3) Auch dies ist wieder ein Aspekt der construct-Validität, hier der 'Gültigkeit' der Auswertung : denn geprüft wird, ob das Auswertungsverfahren (in seinen Ergebnissen) bestimmten Hypothesen standhält. Ein solcher Weg aber kann nur eingeschlagen werden, solange es sich um 'Forschung' handelt und kann nicht verwechselt werden mit dem Prinzip jener Validität, welche eine Voraussetzung für die diagnostische Verwendung des Verfahrens darstellt.

Gesamtscore von der Größe ≥ 0. In dem Gesamtscore für eine bestimmte Variable drückt sich die Summe aller für die Variable vorgenommenen Signierungen aus. Werden also z.B. bei einer Vp 16 +Antworten als ihrem manifesten Gehalt nach Aggression beinhaltend signiert, ergibt sich für diese Vp ein Gesamtscore von : +Agg=16 . Der negative Gesamtscore bezeichnet entsprechend die Summe aller in den Antworten vorkommenden Ablehnungen der betreffenden Variablen, also z.B. - Agg=7 [1), 2)]. Dabei entstammen weitgehend [3)] die positiven scores den positiven Antworten, die negativen scores den Begründungen zu den abgelehnten items.

Die Vpn unterscheiden sich einmal nach der Höhe der Gesamtscores zu einer bestimmten Variablen (also nach der Anzahl der Antworten, welche diese Variable ihrem sprachlich-manifesten Gehalt nach beinhalten). Sie unterscheiden sich aber auch danach, welche Variablen überhaupt im Testprotokoll ausgemacht werden konnten und somit zugleich nach der Zahl der insgesamt überhaupt vorgenommenen Signierungen. Damit entsteht das Problem, daß Kriterien für eine Entscheidung darüber gewonnen werden müssen, wie diese unterschiedliche 'Ergiebigkeit' der Testprotokolle zu bewerten ist. Handelt es sich um unterschiedliche 'Offenheit' in der Reaktion; verweisen jene Testprotokolle, bei denen nur relativ wenig Signierungen möglich sind, auf Verteidigungsverhalten; besitzen die Vpn eine unterschiedliche Fantasie oder wirken die items der WP teilweise mehr als Auslöserreize für Interessen und Kognitionen, anstatt für jene Bedürfnisse bzw. 'Wünsche', deren Ausdruck wir erwarten? Diesen Fragen konnte im Rahmen dieses Erkundungsversuchs nicht im einzelnen nachgegangen werden. Der durch diese unterschiedliche Ergiebigkeit bedingten Relativität der scores für die einzelnen Variablen mußte jedoch Rechnung getragen werden.

---

1) Die Abkürzungen für eine positive Wahl (+Agg) und die Ablehnung (-Agg) werden im folgenden für alle Variablen verwendet.
2) Eine positive Agg-Antwort wäre z.B.: "dann kann ich die Leute ärgern" (+Agg); eine Ablehnung von Aggression z.B. : "weil ich dann zerstören müßte" (-Agg).
3) Dieser Zusammenhang ist näher unter den 'Formalen Aspekten zur Auswertung der WP' im ANHANG erläutert (s.S. 320)

Wenn die Gesamtzahl der bei den Vpn vorgenommenen Signierungen schwankt, ist es nicht möglich, die scores verschiedener Vpn ihrem absoluten Wert nach zu vergleichen. Wird beispielsweise in einem Testprotokoll nAggression 10 mal signiert, so kommt diesem score (+Agg=10) ein unterschiedliches Gewicht zu, je nachdem ob in dem Testprotokoll i n s g e s a m t etwa 20 oder 40 Signierungen haben durchgeführt werden können. Im einen Falle sind 50% aller Antworten, die unseren Variablen zugeordnet werden können, ihrem manifesten Gehalt nach durch nAgg charakterisiert, im anderen Falle sind es nur 25% aller signierten Antworten. Welches relative Gewicht also soll dem gleichen Agg-score von +Agg=10 in beiden Protokollen zukommen?

Wir gehen davon aus, daß die 11 Variablen, nach welchen wir die Testprotokolle auswerten, den gesamten Ausschnitt aller Motivvariablen, deren Ausdruck überhaupt durch die WP ausgelöst werden kann, erfassen. Bei den Testprotokollen unserer Versuchsgruppe gab es zwar eine Fülle von Antworten, die nicht diesen 11 Variablen zugeordnet werden konnten, doch fanden wir keine Antworten, die ihrem sprachlich manifesten Gehalt nach als Ausdruck einer weiteren M o t i v - variablen hätten bewertet werden müssen.

Wenn es aber zutrifft, daß durch unsere Auswertung nicht Motivvariablen in den Test-Antworten u n berücksichtigt bleiben, welche u.U. besonders charakteristisch und daher ausschlußreich sind für einzelne Vpn, so können wir folgendermaßen vorgehen : wir können an jedes Testprotokoll den 'Maßstab' der ü b e r h a u p t zu erwartenden Variablen anlegen und feststellen, welche Variablen eine Vp zum Ausdruck bringt und mit welcher Häufigkeit sie in den Antworten auftreten. Die Summe aller durchgeführten Signierungen kann nun zum Ausgangspunkt der Bewertung gemacht werden : die scores für die einzelnen Variablen (nAgg etc.) werden relativ zur Gesamtzahl der vorgenommenen Signierungen bewertet und zwar ausgedrückt als Prozentzahl (s.o.). Eine Vp mit insgesamt 2 Aggressions-Antworten kann so einen höheren score für nAgg erhalten als eine Vp mit 4 Agg-Antworten, wenn

nämlich das erstere Testprotokoll sehr viel weniger Signierungen enthält als das letztere. [1] Ein solches Vorgehen wäre nicht zulässig, wenn der benutzte Variablen-Katalog nur einen Teil der in den Testprotokollen vorkommenden Variablen erfaßte, wenn etwa eine Reihe von Motiv-bezogenen Testantworten unberücksichtigt bliebe. In diesem Falle wäre das Bezugssystem 'Summe aller durchgeführten Signierungen' willkürlich; die relativen Gewichtungen der scores für die einzelnen Motivvariablen würden empfindlich schwanken, je nachdem welche Variablen gerade als Bezugssystem gewählt würden.

Freilich ist auch dieses Vorgehen nicht unproblematisch. Wenn wir bei e i n e r Vp die scores verschiedener Variablen miteinander vergleichen, können wir nur Aussagen folgender Art machen : nDom (20%) nimmt im Testprotokoll ein größeres Gewicht ein als nAff (8%), und dieses taucht häufiger auf als nExh (2%). Über die 'tatsächlichen' Relationen zwischen den scores dieser Variablen können wir aus zweierlei Gründen zunächst nichts aussagen : einmal können wir den Test nicht 'interpretieren', wir kennen seine Gültigkeit nicht und wissen z.B. nicht, auf welcher 'Ebene' er funktioniert; wollten wir zum anderen auch nur allgemein folgern, nDom spiele bei der Vp eine 'irgendwie' größere Rolle als nExh, so müßte bereits geprüft sein, daß die Auftretenswahrscheinlichkeit beider Variablen die gleiche ist. Denn andernfalls könnte diese Differenz zwischen den scores von nDom und nExh allein daran liegen, daß die WP für nExh sehr viel weniger Auslöserreize enthält [2]. Zunächst können wir nur eine rein formale Auswertung betreiben, indem wir nicht mehr feststellen als das relative Gewicht einzelner Variablen in einem Test p r o t o k o l l . Über das relative Gewicht der Variablen als Verhaltensdeterminanten der P e r s o n können zunächst keine Aussagen gewonnen werden.

---

1) Ähnlich ging auch Murray (1943) vor, wenn er die scores für eine bestimmte Variable zwar über die (unterschiedlich 'ergiebigen') Geschichten hinweg addierte, jedoch zuvor nach der Länge der Geschichten korrigierte. Allerdings sind - ebenso wie die bloße Wortanzahl von Geschichten - die 'überhaupt möglichen Signierungen' kein absoluter Maßstab, sondern bereits begrenzt durch das von außen herangetragene Variablensystem.

2) vgl. dazu die ausführliche Diskussion S.93ff und S.123.

Auf der Grundlage einer solchen Auswertung ist es dann auch in einem ebenso rein formalen Sinne möglich, die Testprotokolle i n t e r -individuell zu vergleichen. Wir stellen fest : nAgg kommt bei Vp A in 20% der Antworten vor, bei Vp B nur in 10% der Antworten. Wir folgern : im Testprotokoll A kommt nAgg ein relativ größeres Gewicht zu als im Testprotokoll B. Wir machen wiederum keine Aussagen über die P e r s o n e n , sondern führen allein einen rein formalen Vergleich der Test p r o t o k o l l e durch.

Aus diesen Prozentwerten können wir demnach 2 verschiedene Arten von Rangreihen bilden : einmal vergleichen wir i n t r a -individuell das relative Gewicht der einzelnen Variablen, zum anderen vergleichen wir i n t e r -individuell das relative Gewicht e i n e r Variablen. Dafür ein Beispiel :

1.intraindividueller Vergl.

Variable	Hfkts. score	%-score	Rangpl.
Dom	12	30	1
Aut	10	25	2
Exh	7	17,5	3
Aff	5	12,5	4
Agg	3	7,5	5,5
Nur	3	7,5	5,5

2.interindividueller Vergl.

Vpn	%-score f. nAgg	Rangpl.
a	7	5
b	12	4
c	15	3
d	30	1
e	27	2
f	4	6

Für den so beschriebenen inter-individuellen Vergleich ist noch eine Bemerkung anzufügen. Wenn wir - ähnlich Murray (s.Fußnote 1), S.199) - die Anzahl der Signierungen für eine Variable nach der 'Ergiebigkeit' des Testprotokolls korrigieren (%-Zahl), so machen wir auch diese 'Ergiebigkeit' zu einem rein f o r m a l e n Kriterium und sehen von den qualitativen Unterschieden hinsichtlich 'Offenheit' etc. ab (vgl. S. 197). Wir setzen damit voraus, daß diese unterschiedliche Ergiebigkeit selbst keine Rolle spielt : denn wir nehmen implizit an, daß bei e i n e r Vp dasselbe Verhältnis der scores verschiedener Variablen zueinander dem Testprotokoll zu entnehmen ist, unabhängig davon, ob sie mit 'großer Ergiebigkeit' oder mit 'geringer Ergiebigkeit' reagierte.

Sind bei einer Vp, deren Testantworten 'unergiebig' waren, z.B. 10% der überhaupt signierten Antworten mit nAgg signiert worden, so nehmen wir implizit an, daß dieselbe Vp - würde sie, aus welchen Gründen auch immer, ausführlicher reagieren - den gleichen (relativen) score für nAgg erhalten würde.

Es sei noch einmal betont, daß wir hier ein rein formales Auswertungssystem sozusagen 'probeweise' einführen. Wir wissen nicht, wie die WP 'tatsächlich' auszuwerten ist, wir kennen nicht ihre Gültigkeit. Wir schlagen, für den Zweck der Analyse dieses Verfahrens, ein Auswertungssystem vor : wie weit es ertragreich ist, können wir erst im experimentellen Teil beurteilen.

Schließlich kann noch eine andere Art der Datenverarbeitung hier vorgenommen werden, nämlich die Kombination der scores v e r s c h i e d e n e r Variablen bei e i n e r Vp. So kann einmal in Anlehnung an die Arbeiten von Mussen und Naylor (1954), aber vor allem von Lesser (1958,1959a,b) der Quotient gebildet werden zwischen Aggressions-Ablehnungen und positiven Aggressions-Antworten (-Agg/+Agg). Ferner lassen sich inhaltlich zusammenhängende Variablen additiv zusammenfassen : so kann etwa der Ausdruck an nExhibition repräsentiert werden durch die Differenz von +Exh und -Exh; oder es können die -Rej-scores den +Aff-scores hinzuaddiert werden, als Ausdruck positiver Ausrichtung auf nAff insgesamt. Die positive Antwort z.B. "dann bin ich nicht so allein" und die negative Begründung "da wäre ich zu allein" betreffen inhaltlich dasselbe und werden nur aufgrund der Bedingungen der items in dieser unterschiedlichen Richtung (+Aff bzw. -Rej) evoziert. [1)]

---

1) Im Fall von nAgg wird ein Quotient gebildet, weil es sich um gegenläufige Tendenzen handelt. Im Fall der Datenkombination von +Aff und -Rej jedoch handelt es sich im Prinzip um gegeneinander austauschbare Signierungen, wie das Beispiel zeigt, so daß hier eine Addition angezeigt ist. Für die Bildung eines Quotienten bei +Exh/-Exh taucht -Exh zu selten auf : bei der Bildung von Summen kann wohl ein Summand, beim Quotienten jedoch nicht einer der Werte fehlen.

In welcher Weise die hier beschriebenen Wunschproben-scores nun weiterverarbeitet bzw. anderen Daten zugeordnet werden, wird erst im folgenden Kapitel zu beschreiben sein.

b) Die Auswertung des Q_s-sort-Verfahrens

Das Datenmaterial des Q_s-sort-Verfahrens ist - von den inhärenten methodologischen Problemen abgesehen - einem Zugriff sehr viel direkter gegeben : die Vpn quantifizieren ja ihre Antworten s e l b s t . Es ist nun hier zu prüfen, welche Möglichkeiten einer Weiterverarbeitung diese direkt quantitativ gegebenen 'Urdaten' bieten. Bei der Darstellung dieser verschiedenen Möglichkeiten wird nicht auf die jeweils zugehörigen methodologischen Probleme hingewiesen werden. Diese wurden in Kap.VI ausführlich erörtert : für eine angemessene Beurteilung der Daten ist diese Diskussion also im Gedächtnis zu behalten.

Es sei hier noch einmal in Erinnerung gebracht, daß jede Vp 88 items (Persönlichkeits-statements; siehe ANHANG) einzuschätzen hat und zwar dies in 3 (real-self, ideal-self, Wunsch-Ich) bzw. 4 (bezieht man auch die Wiederholung des real-self [1]) mit ein) verschiedenen Bezugssystemen. Für jedes einzelne item gibt es also bei jeder Vp in jedem Bezugssystem je einen quantitativen score (Einschätzung auf der 9-Punkte-Skala); jede der 11 Variablen ist dabei durch die Einschätzung von je 8 verschiedenen items repräsentiert. Die Auswertung kann also sowohl nach items, wie (zusammengefaßt) nach Variablen vorgenommen werden.

Das bisher gebräuchlichste Verfahren einer Verarbeitung dieser Daten ist das der Korrelation aller Einschätzungen in einem Bezugssystem mit denen eines anderen Bezugssystems (z.B. real/ideal). Die Gesamt-Korrelations-Koeffizienten, die sich auf diesem Wege ergeben, sind die zwischen real/ideal, real/Wunsch, ideal/Wunsch, real(1)/real(2). Für jede Vp kann also das Verhältnis zwischen je 2 Bezugssystemen auf der Basis a l l e r items festgestellt werden.

Weiter kann dieses Verhältnis zwischen 2 Ebenen auch für

[1] F ü r die Urteils-Bezugssysteme ('Bezugsebenen') werden im folgenden nur noch Abkürzungen verwendet : real(1), ideal, Wunsch und real(2).

jede einzelne V a r i a b l e getrennt analysiert werden : die Einschätzungen der je 8 zu einer Variablen gehörenden items lassen sich gesondert auswerten und auf den verschiedenen Ebenen miteinander vergleichen. Den Grad des Unterschiedes bzw. der Übereinstimmung zweier Ebenen bestimmen wir hier durch einen Differenzen-score (ferner abgekürzt als Diff-score). Für jedes item wird die a l g e b r a i s c h e Differenz [1] zwischen den Einschätzungen zweier Ebenen festgestellt : der Diff-score für eine Variable ergibt sich aus der Summe der für die 8 zugehörigen items berechneten Differenzen. Diese Diff-scores bieten ein Maß sowohl der Größe wie der Richtung des Unterschieds zwischen 2 Ebenen (bezogen auf e i n e Variable) : so kann etwa festgestellt werden, daß für eine Vp nAgg im Verhalten (real) stärker ausgeprägt, als vom Ich-Ideal her zugelassen ist. [2]

Auf eine weitere Kategorie von Daten ist an anderer Stelle schon einmal verwiesen worden, nämlich auf die intra- und inter-individuellen Streuungen, welche eine besondere Rolle spielen für die in dieser Arbeit entwickelten Kriterien einer construct-Validität des Verfahrens (s.S.184ff). Mit dem Begriff der i n t r a -individuellen Streuung ist die Frage gestellt, ob die 8 items einer Variablen relativ homogen, also in einheitlicher Weise auf einem engen Skalenbereich eingeschätzt wurden, oder ob die Einschätzungen sich über die ganze Skala verteilen. Diese Streuung der je 8 zu einer Variablen gehörigen items ist für jede Vp, für jede Variable, wie für jedes Bezugssystem zu berechnen. Da es sich nicht um normalverteilte Daten handelt, ist der Interquartilraum der 8 Einschätzungen als Streuungsmaß zu verwenden.

Weiterhin ist es, wie sich bei der Darstellung des Datenmaterials zeigen wird, ebenfalls von Interesse, für jede Vp die durchschnittliche i n t r a -individuelle Streuung zu

---

1) Es kann nicht die statistische Signifikanz des Unterschiedes berechnet werden, da es sich um einen i n t r a-individuellen Vergleich handelt (vgl. dazu Fußnote 1), S.211).

2) Derartige Relationen werden im folgenden abgekürzt als real > ideal und entsprechend ideal > Wunsch, Wunsch > real etc.

bestimmen, also die durchschnittliche Homogenität bzw. Heterogenität ihrer Einschätzungen der je 8 items zu einer Variablen zum Ausdruck zu bringen. Jede Vp erhält bei 11 Variablen in jedem Bezugssystem 11 Streuungsmaße (Interquartilräume) : es werden diese 11 Werte bei jeder Vp jeweils in ihrem Median [1] zusammengefaßt.

Die i n t e r - individuelle Streuung ist für jedes item getrennt festzustellen; sie wird als Interquartilraum der 34 Einschätzungen zu einem item bestimmt.

Ein anderer Gesichtspunkt für die Auswertung des Datenmaterials besteht darin, die Einschätzung jeder einzelnen V a r i a b l e n , nämlich deren Ausprägungsgrad im real-self, ideal-self und Wunsch-Ich bei e i n e r Vp, durch einen einzigen Wert zu repräsentieren, so daß man z.B. feststellen könnte : Vp A gibt im real-self nAff ein größeres Gewicht als nSuc; oder nDom nimmt im real-self bei Vp B einen höheren Platz ein als bei Vp C; oder aber bei Vp D wird nAgg im Wunsch hoch, im real jedoch gering eingeschätzt. Es wird daher bei jeder Vp für jede der 11 Variablen der Median aus den Einschätzungen der 8 zugehörigen items gebildet. Diese Mediane pro Variable lassen sich also intra-individuell, wie inter-individuell vergleichen. Insbesondere können alle Vpn danach in eine Rangreihe gebracht werden, wie sie in einem bestimmten Bezugssystem (z.B. ideal) eine bestimmte Variable (z.B. nAff) einschätzen. [2]

Schließlich lassen diese Mediane sich noch unter einem weiteren Gesichtspunkt verwerten : mit Hilfe aller Mediane für eine bestimmte Variable in einem bestimmten Bezugssystem (bei 34 Vpn also 34 Mediane) kann festgestellt werden, ob die betreffende Variable im Rahmen dieser Versuchsgruppe relativ einheitlich eingeschätzt wurde (z.B. im ideal nNur durchgehend

1) Da auch diese Daten nicht normalverteilt sind, ist ein non-parametrisches Maß angezeigt und nicht etwa das arithmetische Mittel.
2) vgl. die analoge Bildung von Rangreihen bei der Auswertung der WP (s.S.200).

positiv, nSuc negativ) oder aber sehr heterogen, also verteilt über den ganzen Skalenbereich (z.B. geringe Übereinstimmung hinsichtlich nAff im real). Es ist also hier die Streuung der 34 Mediane bei einer bestimmten Variablen zu berechnen und mit der Streuung einer anderen Variablen zu vergleichen. [1)]

Die Anwendung der hier zusammengestellten Auswertungsgesichtspunkte auf das empirische Datenmaterial und die Bewertung der experimentellen Ergebnisse ist Aufgabe des folgenden Kapitels.

c) Die Auswertung der übrigen Verfahren

Es ist bereits an anderer Stelle gesagt worden (s.S. 124) daß die Auswertung des TAT nicht in diese Arbeit hat aufgenommen werden können. Es erübrigt sich also jeder weitere Bezug auf dieses Verfahren.

Ebenfalls ist schon angedeutet worden (s.S.127), daß auch das Erinnerungsexperiment nicht hat voll ausgeschöpft werden können. Es sei daher nur ein Auswertungsaspekt dargestellt; er ist rein formaler und nicht inhaltlicher Art. Es sei noch einmal in Erinnerung gerufen, daß es hier um die Analyse jener items geht, die von den 66 items der WP eine Woche nach deren Erhebung noch reproduziert werden konnten.

Die von einer Vp erinnerten Wörter werden daraufhin betrachtet, wie sie sich auf die verschiedenen Antworten-Kategorien (+, -, $^{+}/_{-}$, 0) verteilen. So wird beispielsweise festgestellt, daß bei einer Vp 40% der erinnerten Wörter zu den abgelehnten items zählen und 60% der erinnerten Wörter im Testprotokoll positiv beantwortet worden waren. Die Verwendung dieser Daten wird erst bei der Besprechung der Ergebnisse beschrieben (s.S. ff). Eine inhaltliche Betrachtung mußte aus zeitlichen Gründen aufgeschoben werden.

---

1) Auch diese Streuung kann nur als Interquartilraum der 34 Mediane berechnet werden.

VIII. Die Ergebnisse des Experiments

1. Nachweis von Validität und Stabilität des Q_s-sort-Verfahrens

Für die Darstellung der Ergebnisse ist zurückzuverweisen auf die in Kap. VI, Abschnitt 8 (s.S. 184ff) entwickelten Hypothesen und auf die Beschreibung des für deren Bestätigung herangezogenen Datenmaterials (s.S.202).

In H y p o t h e s e I (S. 185) wurde die Erwartung formuliert, daß die inter-individuelle Streuung im real-self (real) [1] am stärksten, im ideal-self (ideal) am geringsten und im Wunsch-Ich (Wunsch) von mittlerer Größe sei. Für eine Prüfung dieser Hypothese bieten sich 2 Möglichkeiten an.

Es können einmal gleichzeitig alle Urteile aller Vpn in einem bestimmten Bezugssystem herangezogen und mit den Urteilen in den anderen Bezugssystemen verglichen werden. Für jedes Bezugssystem müßte also der Grad der Übereinstimmung oder Variabilität der Einschätzungen in einem Wert zusammengefaßt werden können. Dafür bietet sich der Konkordanz-Koeffizient [2] nach Kendall (1948) an, welcher die Übereinstimmung nicht von 2, sondern von beliebig vielen Rangreihen mißt. Bei seiner Verwendung müssen hier die Vpn als 'Urteiler' aufgefaßt werden, welche jeweils 11 Variablen des Q_s-sort's ihrer Rangfolge nach zu beurteilen haben. Im Unterschied zur gebräuchlichen Anwendung des Konk.-Koeff. werden jedoch in diesem Falle die 'Urteile' nicht selbst schon in Form einer Rangreihe abgegeben, sondern vielmehr erst im nachhinein durch den Experimentator in eine Rangreihe transformiert. Eine solche Rangreihe kann für jede Vp aus den Medianen gebildet werden, welche für jede Variable die 8 zugehörigen Einschätzungen (8 items) in ihrem Mittel zusammenfassen (s.S.204). Jede Vp kann also für jedes Bezugs-

---

1) Diese 3 Bezugssysteme oder 'Ebenen' werden im folgenden nur durch ihre Abkürzungen bezeichnet : real, ideal, Wunsch.

2) ferner abgekürzt als Konk-Koeff.

system durch die Rangfolge von 11 Medianen (bei 11 Variablen) repräsentiert werden; und auf der Grundlage dieser Rangreihen kann der Grad der Übereinstimmung der 34 'Urteiler' für jedes Bezugssystem getrennt errechnet werden. Da der Konk-Koeff. den 'Grad der Übereinstimmung der Urteiler' mißt, müßte er,der Hypothese zufolge, für das ideal am größten sein. Es ergeben sich folgende Werte :

Tab. I : Konkordanzkoeffizienten : als Maß der inter-individuellen Übereinstimmung der Einschätzungen in den verschiedenen Bezugssystemen

Bezugssystem	W
ideal-self	o,61
Wunsch-Ich	o,41
real-self(1)	o,15
real-self(2)	o,22

Diese Ergebnisse entsprechen genau den Erwartungen, sieht man davon ab, daß die Einschätzungen im real-self(2) unwesentlich weniger heterogen sind als im real-self(1). Allerdings liefern sie für eine zureichende Beurteilung der Hypothese noch nicht genügend Information, weil für dieses statistische Verfahren bis heute keine Unterschiedsmaße entwickelt worden sind. Die Hypothese wäre erst in befriedigender Weise bestätigt, wenn die Differenzen zwischen den W-Werten als signifikant nachgewiesen werden könnten. Die W-Werte selbst nämlich sind gleichartig hoch-signifikant (1%-Niveau). Es ist daher noch ein weiteres Prüfungsverfahren heranzuziehen.

Hatte das obige Verfahren auch den Vorteil, die durchschnittliche Gruppentendenz in einem einzigen Wert anschaulich zum Ausdruck zu bringen, so sind die Daten nun für jede Variable getrennt zu betrachten. Genauer nämlich ist es, die inter-individuellen Streuungen heranzuziehen, welche jedem einzelnen i t e m in den verschiedenen Bezugssystemen zukommen. Auf diese Weise können die Streuungen (Interquartile von

34 Vpn, s.S.204) für jedes der 88 items getrennt einander gegenübergestellt werden. Sodann aber lassen sich die Streuungen von jeweils den 8 items, welche derselben Variablen zugehören, auch gemeinsam einer statistischen Berechnung unterziehen. Jede Variable ist (bei je 8 items) durch 8 Interquartile repräsentiert, welche nun - nach Wilcoxon [1] - im Paarvergleich je zweier Bezugssysteme auf Größe und Richtung ihrer Differenz hin analysiert werden können. Es wurde dies allerdings nur - beispielhaft - für nAgg berechnet [2].

Tab. II : Unterschiede zwischen den Bezugssystemen i.b. auf die inter-individuelle Streuung der Einschätzungen von nAgg

Bezugssysteme	z	p
real(1) > ideal	2,52	0.01
real(2) > ideal	2,17	0.02
real(1) > Wunsch	2,1	0.02
real(2)/ Wunsch	0,84	/
ideal / Wunsch	1,4	/
real(1)/real(2)	0,56	/

Es ergibt sich, der Hypothese gemäß, daß die inter-individuelle Streuung bei der Einschätzung der Variable nAgg im real sign. größer ist als die inter-individuelle Streuung der Einschätzungen derselben Variablen sowohl im ideal, als auch im Wunsch. Ein sign. Streuungsunterschied zwischen den Einschätzungen im ideal und denen im Wunsch ergibt sich allerdings nicht. - Weiterhin zeigt sich, daß die Streuungsdifferenz zwischen den Bezugssystemen für real(2) weniger stark ausgeprägt ist als für den Vergleich mit real(1). Es

1) Der 'Wilcoxon matched-pairs signed-ranks test' wird hier nach Siegel (1956) verwendet; vgl. dazu Wilcoxon 1945, 1947, 1949.

2) Eine solche Beschränkung auf eine einzige Variable wurde aus zeitlichen Gründen notwendig. Diese Relationen werden vielleicht nicht bei allen Variablen so deutlich ausgeprägt sein, doch können sie wohl der Richtung nach bei allen Variablen erwartet werden.

ist daher als ein Aspekt der Zuverlässigkeit des Verfahrens zusätzlich geprüft worden, ob etwa real(1) und real(2) sich hinsichtlich ihrer inter-individuellen Streuung unterscheiden (für nAgg) : es besteht - für die Brauchbarkeit des Verfahrens entscheidend - keine sign. Differenz zwischen den Streuungen; der z-Wert ist hier, absolut gesehen, der geringste (s.Tab. II).

Somit ist am Beispiel einer einzelnen Variablen bestätigt worden, daß die inter-individuelle Streuung im real sign. am höchsten ist : die Vpn unterscheiden sich also wohl nach ihrem 'Verhalten', so wie sie es einschätzen, weniger jedoch nach ihren Wertmaßstäben für dieses Verhalten. Es ist damit einmal aufgewiesen, daß das Verfahren sich in seiner Aufgabe bewährt, bestimmte theoretisch erwartete Phänomene widerzuspiegeln. Es ist ferner - anhand der je nach Ebene unterschiedlichen Variabilität der Urteile - nachgewiesen worden, daß die Variable 'social desirability' - im Gruppendurchschnitt gesehen - die Einschätzungen im real nicht entscheidend oder jedenfalls sign. geringer beeinflußt hat als die Einschätzungen im ideal : dieser sign. Unterschied aber dürfte als Grundlage für eine Beurteilung der Wirkung von 'social desirability' hinreichend sein (vgl. die ausführliche Diskussion auf S. 180 f, 184). Wenn damit auch nicht im einzelnen geprüft ist, ob etwa eine bestimmte Vp sich hinsichtlich ihrer Aggressivität oder anderen Variablen mehr der idealen Norm als der Realität entsprechend eingeschätzt hat, so ist doch jedenfalls für das V e r f a h r e n nachgewiesen, daß es zwischen der 'social desirability' einer Variablen und deren realer Ausprägung, wie sie in der Selbstwahrnehmung gegeben ist, deutlich zu differenzieren vermag. In diesem Punkt genügt das Verfahren somit insbesondere einem Kriterium, das nach dem Prinzip der construct-validity gewonnen wurde.

Für diesen theoretischen Ansatz, nach welchem die Rolle von 'social desirability' nicht anhand von Mittelwerten, sondern nur auf der Grundlage der Varianz der Urteile beurteilt werden kann, kommt der H y p o t h e s e II (S. 185) die

Funktion der Kontrolle zu. Es wird behauptet, daß auch in unserem Material - nicht anders als in den kritisierten Untersuchungen - sich eine sign. Übereinstimmung der Mittelwerte aller Urteile in real und ideal aufweisen lassen würde, o h n e daß sich damit der Rückschluß auf die störende Wirkung von s-d theoretisch begründen ließe. Es ist also nun auf die Varianz der Urteile zu verzichten. Stattdessen sind die 34 Mediane, welche die durchschnittliche Bewertung einer Variablen getrennt für alle Vpn repräsentieren, für jede einzelne Variable ihrerseits in ihrem Mittel (Median der Mediane) zu repräsentieren [1]. Es sind also für das real und ideal die 11 Mediane (bei 11 Variablen), welche die durchschnittliche Bewertung jeder Variablen durch die Gruppe anzeigen, miteinander zu vergleichen. Die Rangkorrelation zwischen diesen Werten beträgt :

$$r_s = 0{,}843 \quad ; \quad p < 0.01$$

Der üblichen Interpretation folgend, müßte diesen Daten entnommen werden, daß die Selbsteinschätzungen (real) dieser Gruppe im Durchschnitt der 'Wertschätzung' (ideal) der betreffenden Variablen entsprechen. Doch ist anhand der interindividuellen Streuung (Hypothese I) nachgewiesen worden, daß die Beurteilung des eigenen 'tatsächlichen' Verhaltens hinsichtlich der Urteils-Variabilität sign. von der normativen Bewertung desselben Verhaltens (derselben items) abweicht. Die Argumentation über die Rolle der Variablen 'social desirability' (S.179ff) erfährt also eine empirische Bestätigung. Was aber besagt der tatsächlich bestehende korrelative Zusammenhang der Mittelwerte seinerseits? Er besagt nicht mehr, als daß die Rangfolge der durchschnittlichen Bewertung der Variablen in der Selbstbeurteilung (real) der Rangfolge des ihnen durchschnittlich zugesprochenen 'sozialen Wertes' entspricht. Eine Interpretation

1) Es sei noch einmal in Erinnerung gebracht : Der Median einer Variable bei e i n e r bestimmten Vp wird aus deren 8 Einschätzungen zu dieser Variable gebildet. Betrachtet man die 34 Mediane, die bei 34 Vpn so für eine bestimmte Variable (z.B. nAgg im Bezugssystem real-self) gebildet werden, so können auch diese wieder in einem Durchschnittsmaß zusammengefaßt werden (Median der Mediane). Für jede Variable gibt es einen solchen Median (der Mediane) pro Bezugssystem. Die Eigenart der Einschätzung eines bestimmten Bezugssystems läßt sich so charakterisieren durch 11 derartige Mediane (bei 11 Variablen).

solcher Relation zwischen abstraktiv gewonnenen Durchschnittswerten (also nicht zwischen den 'Urteilen' selbst) ist auf S.181 gegeben worden.

Die H y p o t h e s e III (S.186) leitet sich aus der Annahme her, daß bei e i n e m Individuum die Beurteilung der 8 zu einer Variablen gehörigen items im ideal konsistenter sei als im real, da bei einer normativen Bewertung ein kognitiv klareres und weniger situationsgebundenes Urteilsprinzip vorherrsche, während die Einschätzung des offenen Verhaltens stärker nach Situationen zu differenzieren sei. Für das Wunsch-Ich wurde ebenfalls eine stärkere intra-individuelle Konsistenz der Urteile für eine Variable erwartet, da die 'Wünsche' eher kompromißlos sich in einer durchgängigen Bejahung oder Ablehnung von 'Eigenschaften' äußern würden und da - per definitionem (s. die Instruktion, ANHANG S. 337) - hier am wenigsten eine differenzierte Situationsanpassung zu leisten ist.

Bei jeder Vp kann für jede Variable ein Index der intra-individuellen Streuung gebildet werden (Interquartilraum der 8 zu einer Variablen gehörigen Einschätzungen, vgl.S.203), so daß jede Vp 11 Streuungsindizes pro Bezugssystem erhält. Es verbietet sich jedoch, für jede einzelne Vp einen Vergleich dieser 11 Streuungsindizes im einen und im anderen Bezugssystem durchzuführen [1]. Daher fassen wir diese 11 Streuungsindizes in ihrem Median zusammen, als dem 'Gesamtindex der intra-individuellen Streuung (vgl.S.204), so daß wir für alle Vpn auf der Grundlage dieser Gesamtindizes prüfen können, ob die intra-individuellen Streuungen im real

1) Statistische Signifikanzprüfung setzt voraus, daß man zu der vorliegenden Datenstichprobe eine Grundgesamtheit denken kann, aus welcher eine beliebige Zahl von Zufallsstichproben entnommen werden kann : es wird geprüft, ob die vorliegende Stichprobe der Zufallsverteilung jener Stichproben zuzurechnen ist, welche der Grundgesamtheit, für welche H_0 gilt, zugehören. Für den intra-individuellen Vergleich ist jedoch dieser Gedanke einer Zufallsverteilung ohne Sinn, da nicht eine beliebige Anzahl von Datenstichproben gewonnen und ein diesen Stichproben gemeinsames Universum gedacht werden kann. Welche Daten-samples wären z.B. zu korrelieren, damit man eine Zufallsverteilung aller r's etwa für den Vergleich von real(1) und real(2) bei einer einzelnen Vp gewinnen kann? Es gibt nur eine Grundgesamtheit von 'Personen' und entsprechend nur Zufallsstichproben von 'Personen', nicht von items (Tests etc.). Eine sehr klare Diskussion dieses Problems findet sich bei Loevinger (1965).

und ideal verschiedenen Grundgesamtheiten zugerechnet werden können. Für diese Unterschiedsprüfung bei abhängigen Stichproben ist wiederum der Wilcoxon-Test angezeigt (a.a.O.). Die in diesem Fall sehr aufwendige Berechnung haben wir dadurch abgekürzt, daß wir nicht die gesamte Versuchsgruppe, sondern nur eine Zufallsstichprobe von 9 Vpn [1] herangezogen haben. Es ergeben sich folgende Relationen :

Tab. III : Unterschiede zwischen den Bezugssystemen i.b. auf die intra-individuelle Streuung

Bezugssysteme	z	p
real(1) > ideal	1,95	0.05
real(1) > Wunsch	1,72	0.05
real(2) / ideal	1,53	/
real(2) / Wunsch	1,18	/
ideal / Wunsch	0,06	/

Es zeigt sich also, daß die intra-individuelle Streuung der Erwartung entsprechend im real sign. am höchsten ist; für den Vergleich von ideal und Wunsch kann die Nullhypothese, wie erwartet, nicht zurückgewiesen werden.

Allerdings ergibt sich auch hier, daß für real(2) der Streuungsunterschied zwar der Tendenz nach, jedoch nicht sign. vorliegt (vgl. dazu auch Tab. II). Damit deutet sich ein Problem an, auf das schon bei der Herleitung der Hypothesen verwiesen worden war. Hypothese III nämlich kann nur als bestätigt gelten, wenn nachgewiesen ist, daß die Ergebnisse nicht auf einen Positionseffekt zurückgeführt werden können (vgl. S.187). Daraus ergab sich als Mindestforderung - vgl. H y p o t h e s e IV (S.187) - , daß real(1) und real(2) sich nicht hinsichtlich ihrer intra-individuellen Streuungen unterscheiden dürfen.

Es wird daher das Verhältnis der intra-individuellen Streu-

1) Es wurde der Reihenfolge nach jede vierte Vp ausgewählt. Da für die gesamte Versuchsgruppe dieselben Reaktionen erwartet werden und sie unter diesem Aspekt eine vollständig repräsentierte Population mit nur e i n e r (begrenzten) Klassenbestimmung (nämlich Zugehörigkeit zur Versuchsgruppe) darstellt, konnte eine simple random sampling durchgeführt werden. Dieses sample sollte ferner zum Zwecke der Abkürzung des Rechenverfahrens nicht zu groß sein.

ungen in real(1) und real(2) nunmehr für alle Vpn [1] berechnet, was auch hier wieder auf der Basis der jeweiligen individuellen Gesamtstreuungsindizes möglich ist (s.o.). Mit einem Wert von z = 0,94 kann die Nullhypothese nicht zurückgewiesen werden : es läßt sich also kein Unterschied zwischen den intra-individuellen Streuungen in real(1) und real(2) nachweisen. Dieser statistische Zusammenhang kann dahingehend interpretiert werden, daß die jeweils besondere Position einer Erhebung (real(1) bzw. real(2)) im Gesamtverlauf der Untersuchung keine ausschlaggebende Rolle spielt; denn sonst müßte für das an letzter Stelle erhobene real(2) die intra-individuelle Streuung, ebenso wie für ideal und Wunsch, sign. geringer sein als für real(1). [2]

Die H y p o t h e s e n V und VI beziehen sich auf die Frage nach der Stabilität des Verfahrens. Es wurde bereits erörtert, daß hier nicht der direkte Weg der Korrelation zwischen den Test- und Re-Test-Daten möglich ist, sondern daß wir indirekt über eine Verteilung der Korr-Koeff. zwischen real(1) und real(2) vorzugehen haben. Diese Verteilung vergleichen wir mit der Verteilung der Korr-Koeff. zwischen ideal und Wunsch [3].

H y p o t h e s e V beschreibt die Erwartung, daß die Stabilitätskoeffizienten höher sind als die Korr-Koeff.

---

1) Dieser Berechnung liegt eine Gruppe von N=31 zugrunde. Denn das real(2) konnte bei einigen Vpn nicht erhoben werden, da sonst bei ihnen die Sitzungszeit zu weit überschritten worden wäre.

2) Hier mag eingewendet werden, daß die intra-individuelle Streuung in real(2) deshalb derjenigen in real(1) gleich sei, weil - nach einer Woche Abstand - der Übungseffekt, der allein die geringere Streuung in ideal und Wunsch bewirkt haben könnte, wirkungslos sei, daß also für b e i d e , real(1) und real(2), die höhere Streuung eine Folge der Position sei. Dieser Einwand aber wird durch den Hinweis entkräftet, daß die items stark in der Erinnerung haften blieben und daß real(2) zumeist von allen Bezugssystemen am schnellsten gelegt wurde. Zu folgern ist also : trotz Übungseffekt ist die intra-individuelle Streuung in real(2) ebenso hoch wie in real(1).

3) Wir ersparen es uns, diesen Vergleich auch für die Korr-Koeff. zwischen real(1)/ideal und real(1)/Wunsch durchzuführen. Entscheidend ist dabei auch der Gesichtspunkt, daß in den Korr-Koeff. ideal/Wunsch nicht die Werte von real(1) enthalten sind : würde man die Korr-Koeff. zwischen real(1)/real(2) vergleichen mit denjenigen von real(1)/ideal, handelte es sich nicht um unabhängige Daten.

zwischen ideal und Wunsch. Wir bestimmen daher für jede Vp[1] die Differenz zwischen Stabilitäts- und Korr-Koeff., indem wir auf z'-Werte übergehen [2]. Für diese Differenzenverteilung ist die Nullhypothese zu prüfen, daß die Differenzen sign. von einer Zufallsverteilung um Null abweichen. Wir können hier den Wilcoxon-Test für abhängige Verteilungen (a.a.O.) heranziehen oder mit dem t-Test die Mittelwertsdifferenz prüfen, indem wir die Differenzenverteilung mit einer Grundgesamtheit vergleichen, deren Mittelwert m=0 ist.

Nach Wilcoxon erhalten wir einen z-Wert von z = 12,6 : die Nullhypothese kann mit p o.oooo3 zurückgewiesen werden. Mit dem t-Test erhalten wir für df=30 einen t-Wert von t=3,3 : die Nullhypothese kann mit p< o.oo5 zurückgewiesen werden. Somit ist die Differenz zwischen den Stabilitätskoeffizienten und den Korr-Koeff. statistisch gesichert.

H y p o t h e s e VI beschreibt die Erwartung, daß die Verteilungen der Stabilitäts- und Korr-Koeff. sich auch in ihrer Streuung unterscheiden. Wir gehen hier ebenfalls auf z'-Werte über und prüfen die Streuungsdifferenz dieser z'-Verteilungen mit dem F-Test [3]. Mit einem F=1,67 bei df=30 läßt sich die Nullhypothese, daß beide Verteilungen hinsichtlich ihrer Streuung der gleichen Grundgesamtheit angehören, nicht zurückweisen (o.1o> p > o.o5). Der Streuungsunterschied ist also nicht signifikant, wenn auch in der erwarteten Richtung relativ hoch.

Betrachtet man jedoch allein die Größe der Stabilitäts- und Korr-Koeff., kann gefolgert werden, daß beide Verteilungen nicht der gleichen Grundgesamtheit angehören. Dieser statistische Zusammenhang läßt sich dahingehend interpretieren, daß mit dem Vergleich von real(1) und real(2)

1) Auch hier handelt es sich um eine Gruppe von N=31, vgl. Fußnote 1) auf S.213.
2) Fisher (1921) hat gezeigt, daß die z'-Verteilung (eine logarithmische Funktion von r) die Verteilung von r normalisiert.
3) Die Verteilung des Quotienten $s_1^2 : s_2^2$ wurde von Fisher (1936) entdeckt. Die Signifikanzwerte dieses Quotienten wurden von Snedecor (1956) berechnet; er führte zu Ehren Fishers die Bezeichnung F ein.

etwas qualitativ anderes gemessen wird als mit der Korrelation zwischen ideal und Wunsch. Wir fassen die vergleichsweise hohe Korrelation zwischen real(1)/real(2) als Zeichen einer genügenden Stabilität des Verfahrens auf [1]. Die vergleichsweise geringe Korrelation zwischen ideal und Wunsch werten wir als Hinweis darauf, daß mit ideal und Wunsch tatsächlich 'verschiedene Dimensionen' gemessen werden.

Diese Tatsache, daß das Verfahren unterschiedliche Grade der Korrelation hervorbringt, je nachdem welcherart Daten man miteinander vergleicht, führt uns noch zu einem weiteren Schluß : wir sehen uns berechtigt, dem Grad der Korrelation zwischen 2 Ebenen überhaupt eine Bedeutung beizumessen. Zwar wissen wir nicht, was ein bestimmter Grad der Korrelation z.B. zwischen real und Wunsch 'tatsächlich' indiziert, da die Gültigkeit des Q_s-sort-Verfahrens ungeprüft blieb. Doch sehen wir uns berechtigt, anhand der Daten dieses Verfahrens Hypothesen zu prüfen, welche Annahmen über die Rolle der Korrelation zwischen 2 Ebenen enthalten.

---

1) Im Rahmen eines Erkundungsversuchs können wir nicht im vorhinein festlegen, welchen Anforderungen an die 'Stabilität' das verwendete Verfahren genügen soll. Wir wissen genau genommen nicht, w a s das Q-sort-Verfahren mißt; also können wir an dieses Verfahren nicht Bedingungen stellen darüber, w i e g e n a u es messen soll. Es muß für diese Untersuchung jedoch gewährleistet sein, daß die U r t e i l e hinreichend konstant sind : hier muß der Hinweis genügen, daß die Urteile in real(1) und real(2) stärker übereinstimmen als die Urteile in ideal und Wunsch. Wir messen also - der exakten Bedeutung des Begriffes nach - nicht die 'Stabilität' des Verfahrens, sondern die Konstanz der Urteile für die einzelne Vp. Entsprechend benutzen wir den Terminus 'Stabilitätskoeffizient' nicht im Wortsinne, sondern nur zur Vereinfachung der Darstellung (vgl.S.188). - Es sei angefügt, daß in diesem Rahmen nicht die Probleme, die mit der Bestimmung der Stabilität eines diagnostischen Instruments bzw. mit der Re-Test-Methode verknüpft sind, im einzelnen diskutiert werden sollen. Für eine kritische Diskussion des Stabilitätsbegriffs, des allgemeineren Begriffs der Reliabilität und des zugehörigen mathematischen Denkmodells sei auf Hörmann (1964b) verwiesen.

Diese in den relativ hohen Stabilitätskoeffizienten ausgedrückte zeitliche Stabilität der Einschätzungen ist noch innerhalb eines weiteren Zusammenhangs von Interesse. Die Bestätigung der H y p o t h e s e VII (s.S. 189) kann als Hinweis darauf gewertet werden, daß die theoretisch erwartete relativ hohe intra-individuelle Varianz in der Beurteilung des eigenen Verhaltens (real(1)) nicht ausschließlich auf einen 'Mangel an kognitiver Klarheit' zurückzuführen ist. Die Annahme von 'kognitiver Klarheit' gab ja den Ausschlag für die Hypothese, daß die intra-individuelle Varianz in ideal und Wunsch geringer sei als in real (s.S. 186). Die relativ höhere intra-individuelle Streuung der Urteile im real sollte jedoch nicht entsprechend durch einen 'Mangel an kognitiver Klarheit' bedingt sein, sondern - im Gegensatz zu ideal und Wunsch - durch die Situationsgebundenheit des zu beurteilenden Verhaltens. 'Mangel an kognitiver Klarheit' müßte gepaart sein mit einer geringen zeitlichen Stabilität der Urteile. Entsprechend müßte eine geringe intra-individuelle Streuung gepaart sein mit 'hoher kognitiver Klarheit' und demzufolge mit hoher zeitlicher Stabilität. Es ist also zu prüfen, ob zwischen den 'Gesamt-Indizes der intra-individuellen Streuung' (vgl.S.212 und S.204) und den Korr-Koeff. zwischen real(1) und real(2) ein korrelativer Zusammenhang besteht. Die Rangkorrelation ergibt bei N=31 (siehe Fußnote 1),S.213) einen Wert von r_s=o,o56 : die Nullhypothese kann also nicht zurückgewiesen werden. Auch Vpn mit starker intra-individueller Streuung im real(1) können eine hohe zeitliche Stabilität zwischen real(1) und real(2) aufweisen.

Es konnten also alle 5 Hypothesen, welche nach dem Prinzip der construct-validity hergeleitet worden waren, bestätigt werden. Diese Hypothesen wurden aus den Definitionen jener theoretischen Realität deduziert, welche durch das Q_s-sort operational repräsentiert werden sollte. Somit können die Ergebnisse als hinreichende Grundlage dafür bewertet werden, daß die mit diesem Verfahren gewonnenen empirischen Daten sich so interpretieren lassen, wie es in Kap. VI erarbeitet wurde. Eine weitere Bewährung dieses Verfahrens kann - wieder im Sinne einer construct-validity - erfolgen durch den Vergleich der Daten verschiedener Verfahren untereinander, wie es die Aufgabe des nächsten Kapitels ist.

2. Aufweis verschiedener Funktionsebenen bei der Wunschprobe

a) Vorbemerkungen

Im Verlaufe der theoretischen Erörterungen ist immer wieder darauf verwiesen worden, daß es sich bei dieser Untersuchung um nicht mehr als ein Erkundungsexperiment handeln kann. Das Ebenen-Konzept bietet - das sei der Diskussion unserer Ergebnisse noch einmal vorausgeschickt - keine Theorie, aus welcher exakte Hypothesen deduzierbar wären; sondern verschiedene allgemeine Fragestellungen dienen als Leitgesichtspunkte für die Auswertung des empirischen Datenmaterials

Die Untersuchung - auch das sei noch einmal wiederholt - zielt darauf ab, T r e n n k r i t e r i e n zu gewinnen, welche als Ansatzpunkte für eine differentielle Validierung der hier beispielhaft analysierten WP dienen können. Diese Trennkriterien haben die Funktion, Gruppen zu konstituieren (und zu differenzieren), für welche sich systematische Beziehungen zwischen Index und Indiziertem nachweisen lassen. So ist beispielsweise festzustellen, ob sich anhand eines bestimmten Kriteriums eine Gruppe konstituieren läßt, für welche eine Beziehung wie die folgende gilt : der Motivausdruck in der WP differenziert diese Vpn danach, wie sie sich für die betreffende Variable in ihrem Verhalten einschätzen.

Diese Trennkriterien sollen mit Hilfe des Ebenen-Konzeptes formuliert werden. Es kann in dieser Untersuchung jedoch nicht darum gehen, jene empirischen Grundlagen zu liefern, auf welchen direkt eine differentielle Validierung des Verfahrens aufbauen könnte. Vielmehr kann nur aufgezeigt werden, d a ß - z.B. bei der WP - mit einer differentiellen Validität gerechnet werden muß; und es kann deutlich gemacht werden, daß es aufschlußreich ist, Testreaktionen unter dem Gesichtspunkt der 'Ebene' zu analysieren. Das bedeutet : wir können zunächst nur allgemein aufzeigen, daß unsere Fragestellung einen heuristischen Wert hat. Wie weit die Ergebnisse über die spezielle Versuchsanordnung hinaus verallgemeinert werden können, wird im Schlußkapitel noch einmal zu untersuchen sein.

An dieser Stelle ist zunächst auf ein statistisches Problem einzugehen, das die Auswertung der Daten im einzelnen betrifft und dadurch bedingt ist, daß der Versuchsplan nicht an fest umrissenen Hypothesen ausgerichtet ist. Die Auswertung ist vielmehr folgendermaßen charakterisiert : eine Vielzahl von Datenpaaren wird sozusagen 'probeweise' daraufhin analysiert, ob sich irgendwelche systematischen Beziehungen zwischen Selbsteinschätzung (den 'Ebenen') und Testreaktion (WP) aufweisen lassen, ein Vorgehen, das Wilkinson (1951) sehr anschaulich, aber irreführend (s.u.) als 'blind analysis type of experiment' bezeichnet. Solches Vorgehen hat zur Folge, daß - wie nachstehend bei der Besprechung der Ergebnisse im einzelnen deutlich werden wird - allein für die Überprüfung einer einzelnen Frage teilweise mehrere gleichartige Signifikanztests erforderlich sind. Dasselbe statistische Prüfverfahren wird also an verschiedenen Datenstichproben wiederholt. Damit ist das Problem z u f ä l l i g signifikanter Ergebnisse gegeben. Entnimmt man einer Grundgesamtheit, für welche die Nullhypothese (H_0) gilt, eine Reihe von Stichproben, so ist mit einer gewissen Wahrscheinlichkeit damit zu rechnen, daß bei einzelnen Stichproben Werte festzustellen sind, welche signifikant von Null abweichen : bei 100 äquivalenten oder (wie Lienert (1962) es nennt) 'homologen' Stichproben aus einer Grundgesamtheit, für welche (H_0) gilt, ist - ex definitione - zufallsmäßig bei 5 Stichproben eine Signifikanz auf dem 5%-Niveau zu erwarten. Mit welcher Wahrscheinlichkeit aber ist ein zufallsbedingt signifikanter Wert z.B. bei 3 homologen Stichproben zu erwarten?

Die Auftretenswahrscheinlichkeit zufallsbedingt signifikanter Ergebnisse wird bei verschiedenen Autoren [1] mit dem Binom $(p - q)^n$ [2] angegeben. Jones und Fiske (1953) verweisen jedoch

1) vgl. Wilkinson 1951; Baker 1952; Brozek, Tiede 1952; Jones, Fiske 1953; Sakoda et.al. 1954; Lienert 1962. Bei diesen Autoren wird allerdings nicht angegeben, durch wen dieses Binomial-Modell ursprünglich entwickelt und eingeführt wurde.

2) Hier bedeutet p = das Signifikanzniveau der auf Zufallsbedingtheit zu prüfenden Werte, q = 1 - p und N = die Anzahl der durchgeführten (homologen) Signifikanzprüfungen.

auf die Schwächen dieses Modells und geben dem chi-Quadrat-Modell [1] den Vorrang, bei welchem nicht nur, wie bei dem Binomial-Modell, die signifikanten, auf ihre Stichhaltigkeit zu prüfenden Werte, sondern die Werte aller untersuchten Stichproben eingehen. Beide Modelle sind nach Jones und Fiske nur anwendbar auf unabhängige Stichproben, d.h. Zufallsstichproben derselben Grundgesamtheit. Wir werden daher zu untersuchen haben, ob dieser Fall in unserem Material vorliegt.

Dieselben Autoren führen ein weiteres Modell an, um Zufallssignifikanzen bei abhängigen Stichproben, also verschiedenen Datenstichproben derselben Vpn, zu bestimmen. Als Beispiel derartiger Stichproben wird der Fall beschrieben, daß bei 2 verschiedenen samples (z.B. Extremgruppenvergleich) anhand von k verschiedenen Messungen geprüft wird, ob sie verschiedenen Grundgesamtheiten zuzurechnen sind. Es wird eine von Hotelling (1931) entwickelte Statistik (T^2) verwendet, welche auf der Grundlage von k Mittelwertsdifferenzen entscheiden läßt, ob die fraglichen 2 Stichproben signifikant voneinander verschieden sind.

Wenn wir die Stichhaltigkeit unserer Ergebnisse, die wir auf dem Wege der 'blind analysis' gewinnen, beurteilen wollen, stehen wir vor zwei Aufgaben : wir haben zu entscheiden, welches der oben genannten Verfahren anzuwenden ist, und wir haben - diese Frage ist der ersteren vorgeordnet - festzustellen, ob bei der Vielzahl der für einen Auswertungsgang verwendeten Stichproben der Tatbestand der 'Homologie' überhaupt gegeben ist. Diese Frage kann nicht nach allgemeinen bzw. formalen Kriterien entschieden werden : sie ist allein von der Eigenart des Datenmaterials und von der Eigenart des theoretischen Ansatzes her zu beantworten, welchem die Daten zugeordnet sind.

---

1) Dieses Modell wurde 1933 von K.Pearson entwickelt. Es liefert eine 'compound probability' (Baker,1952) auf der Grundlage aller zu einer Hypothese unabhängig voneinander durchgeführten Messungen. Hier wird nicht festgestellt, ob die erhaltenen sign. Abweichungen von H_0 zufallsbedingt sind, sondern mit welcher Wahrscheinlichkeit H_0 zurückgewiesen werden kann, wenn man alle durchgeführten Messungen gleichzeitig berücksichtigt :
$$\chi^2 = -2\log_e(p_1\ p_2\ p_3\ \ldots p_k).$$

Es ist daher von Nutzen, sich noch einmal genau zu vergegenwärtigen, welcher Tatbestand mit dem Begriff 'homologe Stichproben' bezeichnet wird. Die Verwendung eines der oben angegebenen Verfahren zur Prüfung der 'Stichhaltigkeit' signifikanter Werte setzt folgenden Gedankengang voraus : Man geht von der Nullhypothese aus, daß die allen Stichproben gemeinsame Grundgesamtheit beispielsweise keine Korrelation zwischen zwei bestimmten Merkmalen aufweist, und man prüft die Gegenhypothese, daß es sich um eine Grundgesamtheit handle, für welche die zwei betreffenden Merkmale korrelieren. Die Nullhypothese kann nur zurückgewiesen werden, wenn die erhaltenen sign. Korrelationen bei der gegebenen Anzahl homologer Stichproben nicht als nur zufallsbedingt (einer Zufallsverteilung um r = 0 zugehörig) bewertet werden müssen.

Diese Überlegung präzisiert den bei Wilkinson höchst mißverständlichen Begriff der 'blind analysis' (s.o.). Es sind nicht unbesehen alle möglichen, dem Datenmaterial nach scheinbar gleichartigen Korrelationen als homolog anzusehen; nicht das mangels Hypothesen ungerichtete 'Suchen' nach irgendwelchen Zusammenhängen stellt das Problem der Homologie. Sondern es handelt sich um den Vergleich mehrerer Zufallsstichproben, für welche f e s t s t e h t , daß sie - i.b. auf das theoretische Problem - derselben Grundgesamtheit angehören; d.h., es werden an mehreren Stichproben Messungen zur gleichen Hypothese durchgeführt ('replicated tests of a single hypothesis', vgl. Ryan, 1959), oder aber es werden - ebenfalls zur Prüfung e i n e r Hypothese - mehrere Messungen bei denselben Vpn durchgeführt. Der Tatbestand der Homologie ist also Voraussetzung dafür, daß die Stichhaltigkeit der Ergebnisse in der beschriebenen Weise geprüft werden kann. Die Frage der Homologie von Stichproben ist nicht G e g e n s t a n d der Prüfung.

In bezug auf unser Datenmaterial werden wir zu dem Schluß kommen, daß der Auswertung weitgehend nicht homologe Stichproben zugrunde liegen, - obwohl die Charakteristik 'blind analysis type of experiment' auf unser Vorgehen zutrifft. In einigen Fällen ist die Entscheidung allerdings schwierig, ob bestimmte Datenstichproben als homolog zu bewerten sind oder

nicht. Zumeist aber läßt sich zeigen, daß derselbe Signifikanztest nicht zur wiederholten Prüfung derselben Hypothese angewendet wird. Es ist eine Eigenart unseres Materials, daß wir bei derselben Stichprobe von Vpn nicht für alle need-Kategorien dieselben Relationen etwa zwischen Selbsteinschätzung und WP-Reaktion erwarten; d.h. wir erwarten nicht, daß die k Messungen, die für eine Stichprobe durchgeführt werden, derselben Grundgesamtheit angehören. [1] Indessen haben wir - es handelt sich ja um ein Erkundungsexperiment - keine präzisen Hypothesen darüber, in welcher Richtung sich diese Messungen unterscheiden. Insbesondere in den Fällen, in denen die Anzahl der erhaltenen sign. Werte besonders gering ist, werden wir die Ergebnisse mit Vorbehalt zu interpretieren haben : denn wir können nicht beweisen, daß es sich n i c h t um zufallsbedingte Signifikanzen handelt.

Um das Verständnis der im folgenden vorzunehmenden Diskussion der Ergebnisse zu erleichtern, wollen wir schon an dieser Stelle die wichtigsten Formen der Auswertung kurz skizzieren.

---

1) Wir setzen damit - ohne vorhergehende Validierung - voraus, daß die Q-sort-items etwa zur Variable nNur 'etwas anderes messen' als die items zur Variable nDom etc.; und wir setzen voraus, daß diese Messungen als unabhängig voneinander bewertet werden können, daß die Einschätzung von nDom also nicht 'beeinflußt' ist durch die Einschätzung der Variablen nNur. Daß diese Unabhängigkeit in einem strengen Sinne nicht besteht, haben wir im EXKURS, bei der Diskussion der 'forced method' aufgezeigt und dadurch erläutert, daß die Q-sort-Skala als Ordinal-Skala aufzufassen ist : mit der forced method ist verknüpft, daß es nur eine begrenzte Anzahl von Freiheitsgraden in der Beantwortung der items gibt. Wir glauben aber, daß man im Rahmen eines ersten Erkundungsversuchs von diesem methodischen Problem absehen kann, da es die Aufgabe einer solchen Untersuchung ist, nur die Richtung anzugeben, in welcher präzisere Experimente zu folgen haben. Ferner wäre für diese Frage der Unabhängigkeit der Messungen - abgesehen von den Schwächen der Methode - theoretisch zunächst zu klären, ob die Variablen nAgg und nDom oder die Variablen nAff und nNur tatsächlich unabhängig voneinander sind, d.h. ob sie unabhängig voneinander sind in der Form, wie sie in dieser Arbeit (in Orientierung an Murray) definiert wurden.

Die Haupthypothese beinhaltet, daß für die vorliegende Versuchsgruppe nicht mit einheitlichen Relationen zwischen Testreaktion und den gewissermaßen als Außenkriterium fungierenden Q_s-sort-Daten zu rechnen ist. Wir erwarten vielmehr, daß nur für Untergruppen, die sich anhand verschiedener 'Trennkriterien' bilden lassen, derartige systematische Relationen aufgefunden werden können. Um dieser Hypothese eine erste experimentelle Grundlage zu geben, prüfen wir zunächst die Gegenannahme, daß die WP, je nach Variable, bei dieser Gruppe durchgehend 'auf einer bestimmten Ebene funktioniert'. Wir prüfen, ob die inter-individuelle Differenzierung, die die WP-scores in einer bestimmten need-Kategorie (z.B. nExh) liefern, jener inter-individuellen Differenzierung entspricht, welche den Q_s-sort-Daten auf einer der verschiedenen Ebenen zu entnehmen ist. Auf diesem Wege könnte sich ergeben (was wir nicht erwarten), daß die WP-scores für nExh die 34 Vpn z.B. danach differenzieren, wie sie sich auf der Wunsch-Ebene unterscheiden. Das würde bedeuten, daß die WP für diese Vpn bei der Variablen nExh 'auf der Wunsch-Ebene funktionierte'.

Wenn diese Gegenhypothese sich widerlegen läßt, werden wir gewissermaßen einen größeren Maßstab anlegen : wir werden fragen, ob die WP-scores Vpn differenzieren, bei denen die interessierenden Merkmale (z.B. die Selbsteinschätzung) alternativ verteilt sind. Wir werden jeweils zwei Gruppen von Vpn einander gegenüber stellen, bei denen ein bestimmtes Merkmal (z.B. die Selbsteinschätzung auf einer 'Ebene') hoch bzw. gering ausgeprägt ist, und prüfen, ob diese Gruppen i.b. auf die Testreaktion verschiedenen Grundgesamtheiten zugerechnet werden können. Als derartige Merkmale verwenden wir den Grad der Selbsteinschätzung der verschiedenen Variablen auf den verschiedenen Ebenen, den Grad der Differenz zwischen den Selbsteinschätzungen auf zwei Ebenen (bezogen auf eine Variable) und den Grad der Korrelation zwischen zwei Ebenen (bezogen auf alle Variablen zusammen genommen). Zu fragen ist also, ob Unterschiede in der Selbsteinschätzung etc. zu einem analogen Unterschied in der Testreaktion führen bzw. umgekehrt, ob hohen Unterschieden in der Testreaktion vergleichbare Unterschiede bei einem der untersuchten Merkmale entsprechen.

Wir können somit feststellen, ob die WP überhaupt inter-individuelle Unterschiede hervorbringt; und wir können prüfen, i.b. auf welches Merkmal oder 'auf welcher Dimension' Probanden durch die WP differenziert werden. Differenzieren unterschiedliche WP-scores z.B. für nExh Vpn auf der real-Ebene, auf der ideal- oder der Wunsch-Ebene; oder ist es der Grad der D i f f e r e n z zwischen zwei Ebenen, welcher mit den need-scores in der WP in einen systematischen Zusammenhang gebracht werden kann. [1)]

Insbesondere für die Zwecke einer diagnostischen Verwendung eines Tests wie der WP interessieren jedoch feinere inter-individuelle Unterschiede, als sie durch den Vergleich von Extremgruppen erfaßbar sind. Jene Merkmale, von denen zu erkunden ist, ob sie durch die WP indiziert werden, sind nicht alternativ verteilt : wie es beispielsweise der Fall wäre, wenn diagnostiziert werden sollte, ob Vpn schizoid sind oder nicht. Ein Test ist diagnostisch um so aufschlußreicher, je feinere Differenzierungen er leistet. Damit kommen wir zum Problem der differentiellen Validität und zur Frage der Trennkriterien : Gibt es Gruppen, bei denen sich der Test als empfindlich erweist für jene inter-individuellen Unterschiede, welche er für die gesamte Versuchsgruppe nicht widerspiegelte? Was konstituiert eine solche Untergruppe, bei welcher der Test für die inter-individuellen Unterschiede eines bestimmten Merkmals empfindlich ist, für welche also auch Testnormen gebildet werden könnten?

Diese Fragestellung sei noch einmal durch ein Beispiel aus der Literatur veranschaulicht. Lesser (1958, 1959b) zeigt auf, daß Aggressions-Indizes in TAT-Geschichten (TAT-ähnliche Bilder) danach zu unterscheiden sind, ob die Vpn durch Aggressions-Angst (AA) charakterisiert sind oder nicht. Hier gilt jedoch nicht die einfache Beziehung, daß Vpn ohne AA hohe Agg-Indizes und Vpn mit AA geringe oder gar keine Agg-Indizes aufweisen(wie es der obigen Fragestellung entspräche, die dem Extremgruppenvergleich zugrunde liegt). Vielmehr

1) Wir verwenden hier durchgehend den Mann-Whitney-U-Test, als nonparametrisches Verfahren zur Unterschiedsprüfung bei unabhängigen Stichproben.

stellt Lesser fest, daß für die Gruppe ohne AA, und nur für diese, eine korrelative Beziehung zwischen Testverhalten und Kriterium (soziometrische Einschätzung der verhaltensmäßigen Aggression) besteht. Das heißt : der Test leistet eine interindividuelle Differenzierung nur innerhalb der Gruppe von Vpn ohne AA, nicht für die Vpn mit AA. Trennkriterium ist hier die Aggressions-Angst. Wir gehen analog vor, indem wir bei den verschiedenen Untergruppen prüfen, ob ein k o r r e l a t i v e r Zusammenhang besteht zwischen Test-Index einerseits und Merkmal (Selbsteinschätzung) andererseits. [1] Anders gewendet : wir versuchen jene Kriterien aufzudecken, die derartige Untergruppen, für welche der Test in unterschiedlicher Weise gültig ist, konstituieren.

Abschließend wollen wir noch einige Abkürzungen einführen, welche die Übersicht über die Ergebnisse im folgenden erleichtern sollen.

Es sei noch einmal daran erinnert, daß zustimmende Antworten in der WP durch ein + (+Dom), ablehnende Antworten durch ein - (-Aff) charakterisiert werden. Es ergab sich daraus z.B. der Quotient $^{-Agg}/_{+Agg}$ als Ausdruck des Verhältnisses von den ablehnenden zu den zustimmenden Aggressions-Antworten. Bei Gegenüberstellungen von Daten charakterisieren wir genauer den need-score von z.B. nExh in der WP als Exh WP und die Selbsteinschätzung des eigenen Verhaltens als real-Exh oder die Einschätzung derselben Variablen auf der ideal-Ebene als ideal-Exh.

Ebenso ist schon die Bezeichnung der Differenz zwischen den Einschätzungen einer bestimmten Variablen auf zwei verschiedenen Q_s-sort-Ebenen eingeführt worden (vgl. S.203), z.B. Wunsch-Agg > real-Agg : es bedeutet dies, daß die Einschätzungen von nAgg im Wunsch-Ich höher (positiver) sind als im real-self. Wenn allgemein zu beschreiben ist, daß die Differenz zwischen zwei Ebenen betrachtet wird, kürzen wir ab durch

---

1) Wir verwenden hier durchgehend die Spearman'sche Rank-Korrelation, da alle zu korrelierenden Daten sich nur einer Ordinal-Skala zuordnen lassen.

z w e i Querstriche : real//ideal. Wenn angezeigt werden soll, daß die Differenz zwischen den Einschätzungen zweier Ebenen gering ist (annähernd Null), so wird abgekürzt : real≈ideal.

Die Korrelation von Daten des einen Verfahrens mit den Daten des anderen Verfahrens kennzeichnen wir durch e i n e n Querstrich zwischen den Daten, die in die Korrelation eingehen. Es handelt sich dabei entweder um die Korrelation zweier Ebenen auf der Grundlage aller 88 items (Ges.Korr.Koeff., s.S. 202), etwa die Korrelation zwischen der ideal- und der Wunsch-Ebene : ideal/Wunsch. Oder es werden für eine einzelne Variable (nExh) die WP-Daten mit den Q_s-sort-Daten derselben Vpn korreliert : Exh WP / real-Exh. Dabei ist dann noch gesondert anzugeben, um welche Vpn es sich bei dieser Korrelation handelt.

b) Zur Frage einer durchgängigen Gruppentendenz

Bevor wir uns der speziellen Frage nach Trennkriterien zuwenden, welche jene Gruppen differenzieren, für die die WP (unserer Versuchsanordnung nach) unterschiedlich 'funktioniert', wollen wir die Gegenfrage stellen : gibt es - zusammengenommen für alle Vpn - bei einzelnen Variablen einen Zusammenhang zwischen der Höhe ihres Ausdrucks in der WP einerseits und ihrer Gewichtung auf einer der 3 Ebenen des Q_s-sort's andererseits; d.h. funktioniert die WP für einzelne Variablen konsistent, also gleichartig für die ganze Versuchsgruppe, auf e i n e r Ebene?

> Für jede der 8 Variablen, welche in der WP insgesamt am häufigsten auftreten, werden alle Vpn, für welche die betreffende Variable in der WP signiert werden konnte (hier 29 Vpn), in eine Rangreihe gebracht; diese Rangreihe wird verglichen mit denjenigen Rangfolgen, welche sich aus der subjektiven Gewichtung derselben Variablen im Q_s-sort ergibt, und zwar auf jeder der 3 Urteilsebenen. Es wird also für jede Variable die Rangreihe der Vpn auf der Basis der WP-scores verglichen mit der Rangfolge ihrer Selbsteinschätzungen auf jeder der 3 Bezugsebenen.

Eine signifikante Korrelation ergibt sich nur für die Variable avoid :

29 Vpn : avoid WP / ideal : r_s=+0,419 , $p < 0.05$

Wir haben zunächst zu prüfen, ob es sich hier um eine Zufallssignifikanz handelt :

Da für 8 Variablen das Verhältnis zwischen WP-scores und Selbsteinschätzung auf jeder der 3 Ebenen zu untersuchen ist, werden insgesamt 24 Korrelationen durchgeführt. Diesen Korrelationen liegen durchgehend dieselben Vpn[1] zugrunde. Für die Entscheidung, welche Korrelationen als homolog anzusehen sind, ist hier, wie bei entsprechenden Fragen im folgenden, ein Gesichtspunkt leitend, der sich aus der Bewertung der Q_S-sort-Technik ergibt, einer Bewertung, die kennzeichnend ist für unseren theoretischen Ansatz, und auch in der experimentellen Literatur vorherrscht. Ohne daß Validitätsuntersuchungen darüber vorliegen, was das real-self und ideal-self eines Q-sort-Verfahrens im einzelnen indizieren, wird die theoretische Voraussetzung allgemein anerkannt, daß es sich bei diesen zwei Erhebungen um prinzipiell verschiedene Dimensionen handelt, daß die Daten also zwei verschiedenartigen Daten-Universa zugehörig sind. Diese Voraussetzung erweitern wir auf das von uns eingeführte Wunsch-Ich. Zu dieser Annahme, daß die Daten verschiedener Ebenen verschiedenen Grundgesamtheiten zugerechnet werden können, sehen wir uns insbesondere aufgrund jener Inter-Relationen zwischen den Daten unseres Q-sort-Verfahrens berechtigt, die wir in Abschnitt 1) dieses Kapitels aufgezeigt haben (vgl. S.206ff).

Damit ist eine erste Aufgliederung der 24 Korr-Koeff. gegeben : zu prüfen ist, ob die jeweils 8 Korr-Koeff., die sich bei dem Vergleich der WP-scores mit den Daten einer Ebene (z.B. real) ergeben, als homolog anzusehen sind. Hier würde es sich dann um äquivalente Datenstichproben handeln, wenn durch jede der 8 Korrelationen dieselbe Hypothese geprüft werden sollte (vgl. S. 220), etwa die Annahme, daß für diese Versuchsgruppe die need-scores der WP (welcher need-Kategorie sie auch zugehören mögen) mit den entsprechenden Daten der ideal-Ebene korrelieren. Eine derartige Hypothese ist indessen nicht gegeben : wir erwarten vielmehr, daß für verschiedene need-Kategorien unterschiedliche Relationen zwischen Test-Index und Selbsteinschätzungen nachweisbar sind; denn wir erwarten bei denselben Tests, daß verschiedene Variablen unterschiedlichen 'Funktionsebenen' zuzuordnen sind. Dieses hat - erschwerenderweise - zur Folge, daß wir nicht entscheiden können, ob der erhaltene Wert zufallsbedingt ist oder nicht.

---

1) Dieses gilt mit der Einschränkung, daß nicht bei jeder Variablen alle Vpn einen need-scores in der WP erhalten; im Prinzip handelt es sich aber um 'dieselbe' Stichprobe, nämlich unsere Versuchsgruppe.

Wir können hier nur folgendermaßen argumentieren : wenn unsere Gegenhypothese zutreffen würde, dann hätte sich - im Idealfall - für jede Variable eine sign. Korrelation zwischen den WP-scores und den Daten jeweils einer der 3 Ebenen feststellen lassen müssen. Es hätte sich ergeben müssen, daß die WP etwa für nExh auf der Wunschebene (Exh WP/Wunsch-Exh), für nNur auf der ideal-Ebene (Nur WP/ideal-Nur) usw. 'funktioniert'. Eben dies ist nicht der Fall : die Gegenhypothese ist nicht bestätigt. Von der einen sign. Korrelation zwischen avoid-WP/ideal-avoid können wir nicht darauf schließen, daß dieselbe Relation und nur diese auch für andere Stichproben von Vpn gegeben wäre. Eine inhaltliche Interpretation dieser Beziehung ist daher nur mit Vorbehalt möglich; wir kommen darauf an späterer Stelle zurück. Wir folgern nur allgemein, daß wir für diese - nach soziologischen Gesichtspunkten relativ homogene - Gruppe nicht ein einheitliches Kriterium bestimmen können, welches sich in eine systematische (korrelative) Beziehung zu den Test-scores bringen läßt. Es ist daher nun zu prüfen, ob der Test zwischen Gruppen mit extremer Ausprägung eines Merkmals unterscheidet (dichotomisiertes Merkmal).

c) Zur Frage einer Extremgruppen-Differenzierung

Wie einleitend dargelegt, wollen wir mit diesem Auswertungsgang prüfen, ob die WP eine inter-individuelle Differenzierung leistet, wenn man Gruppen heranzieht, in denen folgende Merkmale alternativ verteilt sind : Selbsteinschätzung auf den verschiedenen Ebenen (betrachtet für einzelne need-Kategorien), Differenzen zwischen den Selbsteinschätzungen auf 2 Ebenen (ebenfalls getrennt betrachtet für einzelne need-Kategorien) und Korrelation zweier Ebenen (d.h. das 'Gefüge der Ebenen' insgesamt gesehen).

Für jedes dieser verschiedenen Merkmale stellen wir Extremgruppenvergleiche an. Die schärfste Isolierung von Extremgruppen liefert der Vergleich der beiden äußersten Quartile einer Verteilung. Bei einer Versuchsgruppe jedoch von N = 34 würden die Untergruppen auf diesem Wege für eine statistische

Berechnung zu klein werden. Um zu erreichen, daß N hinreichend groß und das Merkmal stark genug dichotomisiert ist, orientieren wir uns an der Streuung des jeweils betrachteten Merkmals : bei einer hinreichend großen Streuung wählen wir $N_1=N_2=12$, bei geringerer Streuung $N_1=N_2=10$. [1] Auf diese Weise erhalten wir unabhängige Stichproben. Jedoch erfüllen die Daten nicht die Bedingungen, welche mit dem t-Test verbunden sind (Normalverteiltheit der Daten, Intervallskala) : wir verwenden daher den Mann-Whitney-U-Test [2], ein nonparametrisches Verfahren zur Unterschiedsprüfung bei unabhängigen Stichproben. Untersucht werden die Variablen, die in der WP bei den meisten Vpn signiert werden konnten : nAgg, nDom, nNur, nAff, nAut, nExh, avoid. Für die Variable nAgg ist ein besonderer Hinweis nötig. Den übrigen Variablen entsprechend sind auf Seiten des Q_s-sort's die Daten der 3 Ebenen zu unterscheiden : real-Agg, ideal-Agg und Wunsch-Agg. Im Gegensatz zu den übrigen Variablen aber unterscheiden wir die WP-scores als -Agg, +Agg und $^{-Agg}/_{+Agg}$.

1. Werden Vpn, die sich in ihrer Selbsteinschätzung hinsichtlich einer bestimmten need-Kategorie unterscheiden, entsprechend durch die need-scores der WP differenziert?

   Für die oben genannten Variablen sind also die Vpn mit den höchsten bzw. geringsten real-scores (ideal- und Wunsch-scores) zu vergleichen : Es wird die Nullhypothese geprüft, daß diese Vpn sich in ihrer Reaktion auf die WP nicht unterscheiden, also derselben Grundgesamtheit angehören, bzw. die Gegenhypothese, daß die need-scores der einen Gruppe höher sind als die der anderen Gruppe. Da eine gerichtete Hypothese nicht vorliegt, etwa derart, daß Vpn mit hohen real-scores auch höhere need-scores in der WP erhalten würden, muß hier der strengere Maßstab der zweiseitigen Signifikanzprüfung angelegt werden.- Die Selbsteinschätzung einer Vp wird durch den Median der 8 Einschätzungen zu einer Variablen (bei 8 items pro Variable) wiedergegeben (vgl. S. 204).- Für jede Variable werden - bei 3 Bezugssystemen - 3 Extremgruppenvergleiche durchgeführt.

---

1) Es ist aus rechnerischen Gründen vorteilhaft, $N_1=N_2$ zu wählen.
2) verwendet nach Siegel (1956), vgl. Mann, Whitney (1947).

Wir wollen die Übersicht über die Ergebnisse hier, wie im folgenden durch eine Tabelle erleichtern, aus welcher alle durchgeführten Unterschiedsprüfungen erkenntlich werden : sign. Unterschiede werden durch ein X markiert [1].

Tab. IV : Extreme Selbsteinschätzungen verglichen mit WP-scores

	+Agg	-Agg	-Agg/+Agg	Dom	Aff	Nur	Aut	Exh	avoid
real		X				X			
ideal		X							X
Wunsch							X		

Diese sign. Unterschiede beinhalten folgende Relationen :

Vpn mit geringen real-Agg äußern sign. mehr -Agg WP als Vpn mit hohem real-Agg ($U=31$, $N_1=N_2=12$, $p < 0.02$)

Vpn mit hohem ideal-Agg äußern sign. mehr -Agg WP als Vpn mit geringen ideal-Agg ($U=34$, $N_1=N_2=12$, $p < 0.05$)

Vpn mit hohem real-Nur äußern sign. mehr Nur WP als Vpn mit geringem real-Nur ($U=32$, $N_1=N_2=12$, $p < 0.05$)

Vpn mit hohem Wunsch-Aut äußern sign. mehr Aut WP als Vpn mit geringem Wunsch-Aut ($U=27$, $N_1=N_2=10$, $p < 0.05$ [2])

Vpn mit hohem ideal-avoid äußern sign. mehr avoid WP als Vpn mit geringem ideal-avoid ($U=32$ $N_1=N_2=12$, $p < 0.05$)

---

1) In den Tabellen wird nicht direkt auf die Extremgruppen und nicht auf die Richtung der Differenz verwiesen. Die Reihen in den Tabellen zeigen jeweils an, auf grund welcher Daten die Extremgruppen gebildet wurden; die Kolumnen verweisen auf die Daten, für welche bei diesen Extremgruppen eine Unterschiedsprüfung vorgenommen wurde. Für Tab. IV heißt das z.B. : in der Reihe 'ideal' und der Kolumne nDom wird für die Vpn mit höchsten bzw. geringsten ideal-Dom-scores geprüft, ob eine entsprechende Differenzierung bei den WP-Dom-Daten vorliegt. Die Richtung der sign. Differenzen wird im Anschluß an jede Tabelle erläutert.

2) Hier handelt es sich um eine einseitige Signifikanzprüfung, bei zweiseitiger Prüfung gälte : $0.10 > p > 0.05$. Da hier die entgegengesetzte Beziehung, nach welcher Vpn mit geringem Wunsch-Aut besonders hohe Wp-Aut-scores erhalten würden, kaum zu erwarten ist, lassen wir hier den weniger strengen Maßstab der einseitigen Signifikanzprüfung gelten.

Zunächst ist wieder zu prüfen, ob diese 5 sign. Relationen bei einem Gesamt von 27 U-Tests (s.Tab.) als zufallsbedingt bewertet werden müssen, bzw. ob es sich teilweise um homologe Stichproben handelt.

> Für jede einzelne Variable (z.B. nDom) wird an 3 unabhängigen Stichproben (Vpn mit höchsten bzw. geringsten real-, ideal-, Wunsch-scores) geprüft, ob eine Relation zwischen Selbsteinschätzung und WP nachzuweisen ist. Diese Stichproben sind nicht als homolog zu betrachten : einmal gehen wir davon aus, daß die Daten der verschiedenen 'Ebenen' grundsätzlich verschiedenen Grundgesamtheiten angehören (vgl. S.226); zum anderen besteht für diese 3 zu einer Variablen durchgeführten Unterschiedsprüfungen keine gemeinsame Hypothese, denn wir erwarten nicht, daß sich für die verschiedenen Ebenen gleichartige Relationen ergeben. Eine umgekehrte Betrachtung führt zu einem entsprechenden Ergebnis : mit den Daten einer Ebene (z.B. real) werden (bei 9 verschiedenen need-Kategorien, s.Tab.) 9 verschiedene U-Tests vorgenommen. Es wird auch hier theoretisch nicht angenommen, daß diese Stichproben derselben Grundgesamtheit angehören. Es wird nicht 9mal dieselbe Hypothese geprüft, etwa die Hypothese, daß die need-scores der WP - für alle Variablen gleichartig - die Vpn danach differenzieren, wie sie sich auf der real-Ebene einschätzen.

Es gibt demnach kein Kriterium, nach welchem wir die obigen sign. Werte als zufallsbedingt zurückweisen müßten. Wieweit sie generalisiert werden können, sei erst an späterer Stelle geprüft. Hier wollen wir die Ergebnisse zunächst nur allgemein charakterisieren :

Die scores einzelner need-Kategorien führen zu einer inter-individuellen Differenzierung : es ist also - dieser Nachweis war bisher nicht geliefert worden - ergiebig, die WP nach einzelnen need-Variablen auszuwerten. Diese inter-individuelle Differenzierung kann insbesondere dadurch aufgewiesen werden, daß man das 'Außenkriterium' in Begriffen des Ebenen-Konzeptes formuliert : je nach Variablen werden die Vpn entweder auf der real-Ebene oder auf der ideal- bzw. auf der Wunsch-Ebene differenziert. Es läßt sich daher zunächst allgemein folgern : das Ebenen-Konzept bewährt sich, indem es dazu führt, daß Zusammenhänge zwischen Testreaktion und einem 'Kriterium' aufweisbar werden. - Eine inhaltliche Betrachtung der auf diesem Wege gewonnenen Ergebnisse sei weiter unten vorgenommen.

Wir wollen zunächst der hier aufgedeckten Beziehung zwischen

Test-score und Kriterium genauer nachgehen. Es ist die Gegenfrage zu stellen, sozusagen die 'Probe auf's Exempel' zu machen, ob dieselbe inter-individuelle Differenzierung auch nachgewiesen werden kann, wenn man von den WP-Daten ausgeht.

2. Entspricht den inter-individuellen Unterschieden hinsichtlich der need-scores in der WP eine vergleichbare Differenzierung der Vpn auf einer der 'Ebenen' (Q_s-sort) ?

   Es werden hier die Vpn mit hohen bzw. geringen need-scores für die einzelnen Variablen in der WP einander gegenübergestellt. Es wird geprüft, ob diese Gruppen hinsichtlich ihrer Selbsteinschätzung in real, ideal oder Wunsch derselben Grundgesamtheit angehören oder sich sign. unterscheiden. Für jede need-Kategorie wird also mittels Extremgruppenvergleich festgestellt, ob eine analoge Differenzierung auf einer der 'Ebenen' nachgewiesen werden kann. [1)]

Wir geben auch hier eine Tabelle, in welcher zum Zwecke einer besseren Übersicht nur die sign. Werte markiert werden.

Tab. V : Extremgruppenvergleich : extreme WP-scores, verglichen mit der Selbsteinschätzung

WP	real	ideal	Wunsch
-Agg			
+Agg			
-Agg/+Agg			
Dom			X
Aff			
Nur	X		
Aut			
Exh			X
avoid			

1) Einzelne Hinweise für die statistische Auswertung wurden bereits auf S.227f gegeben.

Die sign. Unterschiede beinhalten folgende Relationen :

Vpn mit hohem Dom WP weisen sign. mehr Wunsch-Dom auf als Vpn mit geringem Dom WP ($U=39$, $N_1=N_2=12$, $p<0.05$ [1]).

Vpn mit hohem Nur WP weisen sign. mehr real-Nur auf als Vpn mit geringem Nur WP ($U=41{,}5$, $N_1=N_2=12$, $p<0.05$).

Vpn mit hohem Exh WP weisen weniger Wunsch-Exh auf als Vpn mit geringem WP Exh ($U=41{,}5$, $N_1=N_2=12$, $p<0.05$).

Das Signifikanzniveau dieser 3 U-Werte ließ sich nur bei einseitiger Prüfung sichern. Allein für die Variable nNur jedoch liegt - auf Grund der in Tabelle IV festgestellten Relation - eine gerichtete Hypothese vor : wenn Vpn mit hohem (geringem) real-Nur hohe (geringe) WP-Nur-scores aufweisen, muß auch die umgekehrte Relation gelten, daß Vpn mit hohen (geringen) WP-Nur-scores sich durch hohe (geringe) real-Nur-scores auszeichnen. Die Relationen, die sich für die Variablen nExh und nDom ergeben, müssen wir als statistisch nicht hinreichend gesichert bewerten.

Wir stehen somit vor dem Problem, daß die Ergebnisse der Tabelle IV einer Gegenprobe nicht standhalten (mit Ausnahme von nNur). Diese Inkongruenz ist folgendermaßen zu erklären.

Wenn Extremgruppen, die auf Grund des Merkmals 'a' gebildet werden, einen entsprechenden Unterschied hinsichtlich eines weiteren Merkmals 'b' aufweisen, so läßt sich diese Relation auch als ein korrelativer Zusammenhang begreifen : denn Vpn, die einen hohen score hinsichtlich Merkmal 'a' aufweisen, erhalten einen hohen score auch hinsichtlich Merkmal 'b', während Vpn mit einem geringen score hinsichtlich Merkmal 'a' einen geringen score hinsichtlich Merkmal 'b' aufweisen. Allein für einen Korrelationskoeffizienten von $r=+1$ bzw. $r=-1$ ist der Grenzfall gegeben, daß die Extremgruppen, die auf Grund des Merkmals 'a' und jene, die auf Grund des Merkmals 'b' gebildet werden, identisch sind. Je geringer die Korrelation beider Merkmale, desto weniger erbringen die Unterschiedsprüfungen auf der Grundlage des einen, bzw. des anderen Merkmals kongruente Ergebnisse. Daß

1) Das Signifikanz-Niveau aller 3 U-Werte ließ sich nur bei einseitiger Prüfung sichern.

aber eine Korrelation beider Merkmale (nämlich WP-scores auf der einen und Selbsteinschätzung auf der anderen Seite) für die gesamte Versuchsgruppe n i c h t gegeben ist, haben wir in Abschnitt b) (vgl. S.225ff) aufgezeigt.

Wir haben daher zu folgern, daß die inter-individuelle Differenzierung durch die WP-scores nur relativ unscharf ist : die Vpn mit den geringsten Wunsch-Aut-scores gehören zu jener Gruppe, die mit einem geringen Ausdruck von nAut auf die WP reagiert; doch gehören dieser Gruppe mit geringem nAut-Ausdruck in der WP auch andere Vpn an, die nicht in derselben Weise durch geringe Aut-scores auf der Wunsch-Ebene charakterisiert sind.

Es ergibt sich somit, daß sich zwar ein Zusammenhang zwischen Selbsteinschätzung und WP-Reaktion für die Versuchsgruppe nachweisen läßt, daß diese Relation jedoch zu unscharf ist, als daß eine eindeutige Zuordnung möglich wäre, wodurch sich der Schluß von dem einen Merkmal auf das andere allein hinreichend fundieren ließe. Somit wird durch diese Ergebnisse die Frage nahegelegt, ob sich einzelne Untergruppen bilden lassen, für welche eine präzisere Zuordnung von Testleistung und Kriterium möglich ist. Die Annahme einer differentiellen Validität des Testverfahrens gewinnt also an Gewicht.

Bevor wir dieses Problem einer differentiellen Validität näher untersuchen, haben wir uns noch weiteren Fragen zuzuwenden, die sich auf das Material der Versuchsgruppe als ganzer beziehen. In der Diskussion des Ebenen-Konzeptes war nicht allein die Erwartung leitend, daß ein projektives Verfahren 'auf verschiedenen Ebenen' funktioniere. Ausschlaggebend war vor allem die Annahme, daß mit dem Begriff des 'Ebenen-Gefüges' der Person eine entscheidende Bedingung der Testreaktion beschrieben werden könne : Vpn mit einer hohen Differenz zwischen real und Wunsch reagieren - so nahmen wir an - anders auf einen projektiven Test als Vpn ohne eine solche Differenz; Vpn, die sich hinsichtlich einer need-Kategorie auf der Wunsch-Ebene höher einschätzen als auf der real-Ebene, reagieren anders auf den Test als Vpn, bei denen für dieselbe Variable das Verhalten höher eingeschätzt wird als

die Wunsch-Ebene. Daraus ergibt sich folgende Frage :

3. Entspricht den inter-individuellen Unterschieden hinsichtlich der Differenz zwischen 2 Ebenen eine vergleichbare Differenzierung der Vpn in ihrer Testreaktion?

Für jede einzelne need-Kategorie (z.B. nDom etc.) werden bei jeder Vp die algebraischen Differenzen (s.S. 203) zwischen den Einschätzungen je zweier Ebenen gebildet, also etwa die Differenz in der Einschätzung von nDom für den Vergleich von real/ideal, real/Wunsch, ideal/Wunsch. Jede Vp erhält für jede einzelne Variable (beim Vergleich von 3 Ebenen) 3 Differenzen-scores. Auch hier werden die Vpn mit hohen bzw. niedrigen Differenzen bei e i n e r Variablen einander gegenübergestellt, also etwa die Vpn, die für nDom eine hohe bzw. geringe Differenz zwischen real//ideal aufweisen etc. Für diesen Extremgruppenvergleich sind 2 Formen der Merkmalsdifferenz zu unterscheiden : entweder sind bei einer bestimmten Variablen die Differenzen z.B. für real//ideal durchgehend positiv, d.h. bei allen Vpn gilt real$>$ ideal bzw. real$\approx$ ideal; oder aber es ergibt sich für den Vergleich zweier Ebenen bei einer Variablen, daß für einzelne Vpn eine Differenz der Richtung real $>$ ideal und für andere Vpn eine Differenz in der Richtung ideal$>$ real (oder die negative Differenz real$<$ ideal) vorliegt. Infolgedessen werden bei diesem Extremgruppenvergleich entweder Vpn einander gegenübergestellt, die eine hohe bzw. eine geringe Differenz zwischen 2 Ebenen aufweisen oder aber Vpn, die sich dadurch unterscheiden, daß im einen Falle eine hohe Differenz der einen Richtung und im anderen Falle eine hohe Differenz der anderen Richtung vorliegt (z.B. real$>$ Wunsch und Wunsch$>$ real). Welche Extremgruppen verglichen werden und von welcher Richtung die sign. Differenzen sind, wird auch hier im Anschluß an die Tabelle angegeben. Es ist die Nullhypothese zu prüfen, daß diese Extremgruppen derselben Grundgesamtheit angehören, bzw. die Gegenhypothese, daß sie sich in ihrer Reaktion auf die WP sign. voneinander unterscheiden.

Auch hier sei zunächst anhand einer Tabelle eine formale Übersicht über die Ergebnisse gegeben :

Tab. VI : Extremgruppenvergleich : extreme Differenzen zwischen den Einschätzungen zweier Ebenen, verglichen mit den WP-scores

Diff	-Agg/+Agg	$^{-Agg}/_{+Agg}$	Dom	Aff	Nur	Aut	Exh	avoid
r//i[1)]								
r//W	X	X						X
r//W			X					

Die sign. Unterschiede beinhalten folgende Relationen :

Vpn mit Wunsch-Agg > real-Agg haben sign. mehr -Agg WP als Vpn mit real-Agg > Wunsch-Agg ($U=32{,}5$, $N_1=N_2=12$, $p < 0.05$)

Vpn mit Wunsch-Agg > real-Agg haben sign. mehr $^{-Agg}/_{+Agg}$ WP als Vpn mit real-Agg > Wunsch-Agg ($U=26$, $N_1=N_2=10$, $p < 0.05$[2)])

Vpn mit real-avoid > Wunsch-avoid haben sign. mehr avoid WP als Vpn mit real-avoid ≈ Wunsch-avoid ($U=17$, $N_1=N_2=10$, $p < 0.02$[1)])

Vpn mit ideal-Nur > real-Nur haben sign. mehr Nur WP als Vpn mit ideal-Nur ≈ real-Nur ($U=17{,}5$, $N_1=N_2=10$, $p < 0.02$)

Vpn mit Wunsch-Dom > ideal-Dom haben sign. mehr Dom WP als Vpn mit ideal-Dom > Wunsch-Dom ($U=21$, $N_1=N_2=10$, $p < 0.05$)

Auch hier ist zunächst zu prüfen, ob die 5 sign. Relationen bei einem Gesamt von 27 U-Tests als zufallsbedingt zurückzuweisen sind; wir haben also zu untersuchen, ob teilweise homologe Stichproben vorliegen.

Für jede Variable (z.B. nExh) wird an 3 verschiedenen Stichproben (nämlich Vpn mit hohen bzw. geringen Differenzen für real//ideal, real//Wunsch, ideal//Wunsch) geprüft, ob sich ein systematischer Zusammenhang zwischen WP-scores und diesen Differenzen nachweisen läßt. Es handelt sich dabei nicht um ein einheitliches Merkmal und nicht um e i n e Hypothese; denn es werden unterschiedliche Relationen erwartet, je nachdem ob die Differenz real//ideal, real//Wunsch etc. betrachtet wird.- Umgekehrt betrachtet

---

1) Die einzelnen Bezugsebenen werden hier, wie bei den weiteren Tabellen durch die Anfangsbuchstaben abgekürzt.

2) Dieses Signifikanz-Niveau ist nur bei einseitiger Prüfung gesichert; hier liegt auf Grund der Relation für -Agg eine gerichtete Hypothese vor.

kommen wir zu demselben Ergebnis : für die Differenz zwischen 2 bestimmten Ebenen wird an 8 verschiedenen Stichproben (bei 8 verschiedenen Variablen [1])) geprüft, ob eine systematische Relation zwischen WP-scores und Grad der Differenz besteht. Wir erwarten jedoch nicht, daß der Grad der Differenz zwischen 2 Ebenen sich bei allen Variablen gleichzeitig auf die WP-Reaktion auswirkt, d.h. wir gehen theoretisch nicht davon aus, daß die Stichproben derselben Grundgesamtheit angehören. Es handelt sich auch in diesen Fällen nicht um 'replicated tests of a single hypothesis' (vgl.S.220). Wir sehen daher die Gruppenunterschiede, die durch die Werte der Tab. VI angezeigt werden, als statistisch gesichert an; der Einwand einer möglichen Zufallsbedingtheit kann auch hier zurückgewiesen werden.

Wir ziehen auch für diese Werte zunächst eine allgemeine Konsequenz; eine inhaltliche Interpretation wird an späterer Stelle vorgenommen.

Es erweist sich hier als aufschlußreich, den Begriff des 'Ebenen-G e f ü g e s ' zur Analyse von Testreaktionen (WP) heranzuziehen : für Vpn, die sich in ihrem Ebenen-Gefüge unterscheiden, lassen sich unterschiedliche Testreaktionen nachweisen. Der Test ist demnach nicht nur empfindlich für Unterschiede in der Selbsteinschätzung auf einer der Ebenen, sondern auch für Unterschiede in den Relationen zwischen 2 Ebenen. Das Konzept des Ebenen-Gefüges bewährt sich in allen 3 Formen, die wir experimentell hergestellt haben : je nach Variable sind die Relationen zwischen real//ideal, real//Wunsch, ideal//Wunsch ausschlaggebend für die Testreaktion.

Auch für diese Schlußfolgerung wollen wir die Gegenprobe machen, indem wir prüfen, ob wir dieselben Ergebnisse erhalten, wenn wir von den Daten der WP ausgehen :

4. Entsprechen den inter-individuellen Differenzen hinsichtlich einzelner need-scores in der WP vergleichbare Unterschiede zwischen den Vpn hinsichtlich ihrer Differenzen-scores.

---

1) Auf Grund der 3 verschiedenen Signierungen von nAgg in der WP lassen sich nur 2 verschiedene Stichproben bilden.

Es sind hier noch einmal die Vpn mit hohen bzw. geringen need-scores in der WP einander gegenüberzustellen (vgl. Tab. V). Es wird geprüft, ob diese Vpn hinsichtlich ihrer Diff.-scores derselben Grundgesamtheit angehören oder sich nach diesem Kriterium sign. unterscheiden. Für jene 5 Variablen, für welche eine derartige Relation zwischen WP-scores und Diff.-scores schon, ausgehend von den Diff.-scores, gefunden wurde (Tab. VI), genügt eine einseitige Signifikanzprüfung, da hier eine gerichtete Hypothese vorliegt.

Es sei wieder anhand einer Tabelle eine formale Übersicht über diesen Auswertungsgang gegeben :

Tab. VII : Extremgruppenvergleich : extreme WP-scores, verglichen mit den Diff.-scores

WP	r//i	r//W	i//W
-Agg			
+Agg			
-Agg/+Agg			
Dom			X
Aff			
Nur	X		
Aut			
Exh			X
avoid		X	

Die sign. Unterschiede beinhalten folgende Relationen :

Vpn mit hohem Dom WP haben sign. größere Differenzen der Form Wunsch > ideal als Vpn mit geringem Dom WP
($U=23{,}5$, $N_1=N_2=12$, $p < 0.01$)

Vpn mit geringem Nur WP haben sign. größere Differenzen der Form ideal > real als Vpn mit hohem Nur WP
($U=35{,}5$, $N_1=N_2=12$, $p < 0.025$)

Vpn mit geringem Exh WP haben sign. größere Differenzen der Form Wunsch > ideal als Vpn mit hohem Exh WP
($U=23$, $N_1=N_2=12$, $p < 0.02$)

Vpn mit hohem avoid WP haben sign. größere Differenzen der Form real > Wunsch als Vpn mit geringem avoid
($U=25{,}5$, $N_1=N_2=12$, $p < 0.01$)

Die Diskussion einer möglichen Äquivalenz der Stichproben und damit einer möglichen Zufallsbedingtheit der sign. Unter-

schiede können wir hier übergehen, da sie denselben Gesichtspunkten folgt, die bei der Bewertung der übrigen Auswertungsschritte leitend waren.

Der Vergleich der Tabellen VI und VII zeigt eine größere Kongruenz zwischen den Ergebnissen als der Vergleich der Tabellen IV und V : für 3 Variablen, nämlich nDom, nNur und avoid führen beide Auswertungsschritte zu denselben Ergebnissen. Jene Vpn, die sich hinsichtlich ihrer WP-Dom-scores unterscheiden, lassen sich auch hinsichtlich ihrer Diff.-scores (Wunsch//ideal) verschiedenen Grundgesamtheiten zuordnen (Tab. VII). Betrachtet man umgekehrt die Vpn, die sich bei derselben Variablen ihren Diff.-scores nach (Wunsch//ideal) unterscheiden, läßt sich entsprechend nachweisen, daß sie in bezug auf ihre WP-Reaktionen verschiedenen Grundgesamtheiten zuzurechnen sind (Tab. VI). Die inter-individuelle Differenzierung ist hier also schärfer als für den Fall, daß man die Selbsteinschätzung auf einer Ebene allein betrachtet (vgl. Tab. IV und V). Das bedeutet zugleich, daß der Zusammenhang zwischen Testleistung und Kriterium für die Diff-scores enger ist als für die Selbsteinschätzung auf einer Ebene. Eine Betrachtung des Ebenen-Gefüges erweist sich hier als heuristisch fruchtbarer als die Betrachtung einer Ebene allein.

Die hohe Übereinstimmung zweier Merkmale hinsichtlich der Extrembereiche der Verteilungen führt zu der Frage, ob sich dieser Zusammenhang auch in Form einer Korrelation beider Merkmale, bezogen auf die gesamte Verteilung, ausdrücken läßt. Wir prüfen daher für die 3 Variablen, für welche sich diese Übereinstimmung besonders deutlich aufzeigen ließ, nämlich für nDom, nNur und avoid, ob sich für die Versuchsgruppe insgesamt ein korrelativer Zusammenhang nachweisen läßt zwischen den need-scores in der WP einerseits und jenen Diff-scores, für welche sich jeweils der Unterschied der Extremgruppen statistisch sichern ließ. Eine solche Korrelation ergibt sich für die gesamte Gruppe nur für nDom. Den Tabellen VI und VII zufolge interessieren hier die Diff-scores für den Vergleich von Wunsch//ideal. Für diese besteht eine

sign. Korrelation mit den WP-Dom-scores :

$$r_s = 0.526 \qquad p < 0.01$$ [1)]

Da in dieser Berechnung positive Differenzen der Form Wunsch-Dom ideal-Dom (hoher Rangplatz) und negative Differenzen der Form Wunsch-Dom ideal-Dom (geringer Rangplatz) in eine gemeinsame Rangreihe gebracht wurden, ist diese Relation folgendermaßen zu interpretieren : je größer der Wunsch ist im Verhältnis zum ideal, desto häufiger wird in der WP mit nDom reagiert;, je geringer aber der Wunsch ist im Verhältnis zum Ich-Ideal, desto geringer ist der Motiv-Ausdruck in der WP. Eine derartige Korrelation zwischen WP-Daten und Selbsteinschätzung ergibt sich nicht, wenn man die Wunsch-Dom-scores oder die ideal-Dom-scores für sich genommen heranzieht; sie ergibt sich nur für die Differenzen zwischen beiden Ebenen.

Durch dieses Ergebnis wird die Behauptung relativiert, daß für die gesamte Versuchsgruppe eine systematische Relation zwischen Testreaktion und Kriterium n i c h t zu erwarten sei. Gleichwohl wird die Annahme, daß derartige systematische Beziehungen vorwiegend für einzelne Untergruppen gelten, durch dieses eine Ergebnis und durch jenes für avoid (vgl.S.225) nicht entkräftet. Von diesen Ausnahmen abgesehen, haben wir bisher keinen k o r r e l a t i v e n Zusammenhang zwischen Testleistung und Kriterium nachweisen können. Nur auf der Grundlage eines solchen korrelativen Zusammenhangs aber lassen sich jene Normen gewinnen, welche eine feinere inter-individuelle Differenzierung ermöglichen, als mit der Unterscheidung von Extremgruppen gegeben ist.

Bevor wir uns der Frage zuwenden, ob sich anhand unseres Materials Untergruppen bilden lassen, bei welchen sich der Test als empfindlich erweist für solche feineren inter-individuellen Unterschiede, haben wir noch einen weiteren Extremgruppenvergleich anzuführen.

Die Q_s-sort-Daten liefern ein weiteres Merkmal, anhand dessen

---

1) Hier genügt eine einseitige Signifikanzprüfung, da eine gerichtete Hypothese vorliegt.

Extremgruppen gegenübergestellt und in ihrer Reaktion auf die WP verglichen werden können : der Grad der Korrelation zwischen den Ebenen insgesamt gesehen (d.h. auf der Grundlage aller 88 items). Es sind diese - wie wir sie zur Abgrenzung gegen die übrigen Korrelationen nennen - 'Gesamt-Korrelations-Koeffizienten' (Ges.Korr.Koeff.), welche in den bisherigen Arbeiten, in welchen die Q_s-Technik verwendet wird, zumeist allein herangezogen werden. Es gibt eine Fülle von Validitätsuntersuchungen, in welchen diesen Ges.Korr.Koeff. - wenn auch auf widersprechende Weise [1] - diagnostische Bedeutung zugesprochen wird.

Wir haben die Gültigkeit unseres Q_s-sort's nicht geprüft und haben demzufolge keine bestimmten Hypothesen über die Rolle dieser Ges.Korr.Koeff.. Es soll nur 'erkundet' werden, ob das in diesen Ges.Korr.Koeff. beschriebene Verhältnis zwischen 2 Ebenen in irgendeiner Weise in Beziehung gebracht werden kann zu den Testreaktionen. Während als 'Kriterium' bisher nur Meßwerte einzelner Variablen dienten, wird also nun gefragt, ob für eine allgemeine Charakteristik ein Zusammenhang mit den need-scores der einzelnen Variablen aufweisbar ist.

5. Entsprechen den inter-individuellen Differenzen hinsichtlich der Ges.Korr.Koeff. Unterschiede in der Testreaktion?

   Es liegen 3 verschiedene Formen von Ges.Korr.Koeff. vor (real/ideal, real/Wunsch, ideal/Wunsch), für welche Extremgruppen zu bilden sind. Für diese Gruppen wird die Nullhypothese geprüft, daß sie sich in ihrer Reaktion auf die WP nicht unterscheiden, bzw. die Gegenhypothese, daß sie verschiedenen Grundgesamtheiten zugerechnet werden können. Die Reaktionen auf die WP werden wieder nach Variablen getrennt betrachtet.

---

1) vgl. hierzu die ausführliche Diskussion im EXKURS, S.172ff.

Wir geben zunächst eine formale Übersicht über den Rechenvorgang :

Tab. VIII : Extremgruppenvergleich : extreme Ges.Korr.Koeff. verglichen mit den WP-scores

Korr.	-Agg	+Agg	-Agg/+Agg	Dom	Aff	Nur	Aut	Exh	avoid
r/i									
r/W					X				
r/W		X							

Die sign. Unterschiede beinhalten folgende Relationen :

Vpn mit hoher Ges.-Korrelation ideal/Wunsch äußern sign. mehr +Agg WP als Vpn mit geringer Ges.-Korrelation ideal/Wunsch ($U=21$, $N_1=N_2=10$, $p<0.05$)

Vpn mit geringer Ges.-Korrelation zwischen real/Wunsch äussern sign. mehr Aff WP als Vpn mit hoher Ges.-Korrelation zwischen real/Wunsch ($U=16,5$, $N_1=N_2=10$, $p<0.02$).

Zunächst ist zu prüfen, ob die Signifikanzen als zufallsbedingt zu bewerten sind.

Dieser Auswertungsschritt unterscheidet sich von jenen, die in den Tabellen IV - VII beschrieben wurden, dadurch, daß alle Unterschiedsprüfungen z.B. für die Korrelation zwischen real/ideal an derselben Stichprobe von Vpn vorgenommen werden. Bei denselben Vpn wird für 9 verschiedene Merkmale, nämlich für 9 need-Kategorien, geprüft, ob sie verschiedenen Grundgesamtheiten zuzurechnen sind. Zwar handelt es sich auch hier nicht im strengen Sinne um 'replicated tests of the same hypothesis'; denn wir erwarten nicht, daß diese Vpn sich bei allen need-Kategorien in derselben Richtung unterscheiden, bzw. daß sie sich überhaupt hinsichtlich jeder Variablen unterscheiden. Wir prüfen nur, o b irgendein Zusammenhang zwischen Ges.Korr.Koeff. und WP-scores besteht. Doch erweist sich hier das Vorgehen einer 'blind analysis' als besonders unzureichend, weil eine eindeutige Entscheidung hinsichtlich der Äquivalenz der Stichproben kaum möglich ist. Geht man allein davon aus, daß 9 gleichartige Signifikanzprüfungen vorgenommen werden, so ist für den Fall, daß es sich um homologe Stichproben handelt, die Wahrscheinlichkeit einer Zufallssignifikanz zu groß, als daß H_0 zurückgewiesen werden könnte. 1) Wir können weder beweisen, daß die Daten verschiedener need-Kategorien verschiedenen Grundgesamtheiten zugehören, noch können wir entscheiden, ob es sich um Zufallssignifikanzen handelt.

1)Handelte es sich um unabhängige homologe Stichproben, so wäre diese Wahrscheinlichkeit für 2 zufallsbedingte Signifikanzen p = 0.0712 (nach Wilkinson, 1951).

Da wir eine klare Entscheidung nicht fällen können, sehen wir von einer Interpretation dieser Werte ab. Wir werden die Ges.Korr.Koeff. jedoch in einem anderen Zusammenhang noch einmal als Kriterium heranziehen (s.u.).

d) Zur Frage einer differentiellen Validität

Wir haben in der Diskussion der bisher dargelegten Ergebnisse festgestellt, daß - im Rahmen unserer Versuchsbedingungen - eine systematische Relation zwischen Test-Index und 'Kriterium' für die gesamte Versuchsgruppe nur in einzelnen Fällen nachzuweisen ist. Als 'Kriterium' dienten dabei zum einen die Selbsteinschätzung hinsichtlich einer need-Kategorie, und zwar betrachtet für jede 'Ebene' getrennt, und zum anderen jene Differenzen, die sich aus dem Vergleich der Einschätzungen zweier Ebenen ergeben. Anhand von Extremgruppenvergleichen ließ sich allgemein zeigen, daß die Relation zwischen Test-Index und diesen Differenzen enger ist als die Relation zwischen Test-Index und Selbsteinschätzung auf nur einer Ebene.

Wenn wir uns jetzt der Frage zuwenden, ob sich für einzelne Untergruppen ein noch engerer Zusammenhang zwischen Test-Index und Kriterium nachweisen läßt, so gehen wir auch hier wieder in Form einer 'blind analysis' vor : denn wir haben keine Hypothesen darüber, welcherart Merkmale derartige Gruppen konstituieren. Wir verwenden probeweise wieder die Kriterien, die sich dem vorliegenden Material entnehmen lassen : Art der Selbsteinschätzung auf einer Ebene, Grad der Differenz zwischen den Einschätzungen zweier Ebenen (betrachtet für einzelne Variablen) und Grad der Korrelation zwischen den Ebenen (betrachtet für alle Variablen). Auch hier werden wieder die extremen Ausprägungsgrade eines Merkmals herangezogen : ist die extreme Ausprägung z.B. von nDom auf der real-Ebene eine Bedingung dafür, daß für diese Vpn die WP einheitlich auf einer Ebene 'funktioniert'; oder wird durch eine hohe Differenz zwischen 2 Ebenen eine Gruppe konstituiert, deren inter-individuelle Unterschiede durch die WP-Indizes widergespiegelt werden; oder führt eine geringe Ges.-Korrelation zwischen 2 Ebenen dazu, daß sich für diese Vpn eine systema-

tische Beziehung zwischen Test-Index und dem Grad der Selbsteinschätzung auf einer der Ebenen aufweisen läßt?

Wir benutzen hier also nicht die Extremgruppenmethode; sondern wir prüfen, ob die extreme Ausprägung eines Merkmals eine Gruppe konstituiert, für welche den inter-individuellen Unterschieden in der Testreaktion entsprechende inter-individuelle Unterschiede eines bestimmten Merkmals zugeordnet werden können. Es gilt jene Gruppen aufzufinden, für welche eine eigene Gruppennorm bestimmbar wäre, für welche sich etwa zeigen ließe, daß die Unterschiede in den Test-scores die Unterschiede im Verhalten wiedergeben. Es sind also Gruppen zu bestimmen, für welche eine k o r r e l a t i v e Beziehung zwischen Test-Index und Kriterium besteht; und es ist zu analysieren, mit welchen Merkmalen sich die WP-Daten in einen korrelativen Zusammenhang bringen lassen. Sollten solche Untergruppen aufgestellt werden können, so wäre - im Rahmen der vorliegenden Versuchsbedingungen - aufgezeigt, daß die WP nicht in derselben Weise für alle Vpn 'valide' ist. Es wäre ein erster Hinweis gegeben auf den Tatbestand einer 'differentiellen Validität'.

Es werden auch hier nur die Variablen, die bei den meisten Vpn signiert werden konnten, untersucht (nAgg, nDom, nAff, nNur, nAut, nExh, avoid); für die Variable nAgg werden, wie oben, die WP-scores unterschieden als -Agg, +Agg, $^{-Agg}/_{+Agg}$.

Zunächst seien die verschiedenen Untersuchungsschritte behandelt; die einzelnen Ergebnisse werden dann in ihrem Zusammenhang interpretiert.

1. Bilden Vpn mit gleichartiger (extremer) Selbsteinschätzung auf einer der Ebenen eine Gruppe, für die sich ein korrelatives Verhältnis zwischen Test-Index und Kriterium nachweisen läßt?

   Für jede Variable werden auf jeder der 3 Ebenen je 2 Extremgruppen, d.h. insgesamt 6 Gruppen gebildet : z.B. die Vpn mit den höchsten (geringsten) Agg-scores in real, ideal und Wunsch [1]. Für diese Gruppen wird geprüft, ob eine Korrelation zwischen dem Test-Index der betreffenden Variable und der Selbsteinschätzung auf einer der Ebenen

1) Für die Größe dieser Gruppen gilt das auf S.228 Gesagte.

aufweisbar ist. Das bedeutet, daß für die 6 Gruppen, die auf der Grundlage der Q-sort-Daten zu einer Variable gebildet werden, je 3 Korrelationen durchgeführt werden (Korrelationen mit den Daten von 3 Ebenen).

Dieses Vorgehen sei auch hier wieder durch eine Tabelle übersichtlich gemacht. Das Material ist allerdings so umfangreich, daß wir eine solche Tabelle nur beispielhaft für e i n e Variable wiedergeben. Eine sign. Relation wird wieder durch ein X markiert. Der Tabelle sind die 6 Untergruppen zu entnehmen und die für jede Gruppe durchgeführten 3 Korrelationen.

Tab. IX : Extreme Selbsteinschätzungen im Q_s-sort als Trennkriterium (Korrelation zwischen WP-scores und Q_s-sort-scores für einzelne Untergruppen)

Selbsteinschätzung		WP/real	WP/ideal	WP/Wunsch
real-Dom:	hoch			
	gering			
ideal-Dom:	hoch			
	gering			X
Wunsch-Dom:	hoch			
	gering			

Dieser Tabelle ist zu entnehmen, daß von den 18 für die Variable nDom durchgeführten Korrelationen nur eine einzige signifikant ist : für die Vpn mit den geringsten ideal-Dom-scores ergibt sich eine sign. Korrelation zwischen den WP-Dom-scores und den Wunsch-Dom-scores. Auch für die übrigen Variablen ergibt sich jeweils nur e i n e derartige Korrelation; für die Variablen nNur und nAff ergibt sich kein sign. Zusammenhang. Die Ergebnisse sind die folgenden :

Vpn mit den geringsten real-Agg-scores :

-Agg/+Agg WP / Wunsch-Agg : r_s=+0,592 , $p < 0.05$ 1)

Vpn mit den geringsten ideal-Dom-scores :

Dom WP / Wunsch-Dom : r_s=+0,619 , $p < 0.05$

Vpn mit den geringsten ideal-Aut-scores :

Aut WP / real-Aut : r_s=+0,541 , $p < 0.05$

Vpn mit den höchsten Wunsch-Exh-scores :

Exh WP / ideal-Exh : r_s=+0,612 , $p < 0.05$

Vpn mit den geringsten real-avoid-scores:

avoid WP / real-avoid : r_s=+0,499 , sign. bei r_s= 0,506

Auch hier ist zunächst wieder der Frage nachzugehen, ob die erhaltenen sign. Korr-Koeff. im Vergleich zu der hohen Gesamtzahl durchgeführter Korrelationen als zufallsbedingt zurückzuweisen sind; es ist also zu klären, ob äquivalente Stichproben vorliegen.

Leitend ist hier, wie in den bisher diskutierten Fällen, die Grundannahme, daß Daten verschiedener Ebenen verschiedenen Grundgesamtheiten angehören (vgl. S. 226). Da wir ferner auch für verschiedene Variablen verschiedenartige Relationen erwarten, können auch die Daten verschiedener need-Kategorien nicht als zur gleichen Grundgesamtheit zugehörig angesehen werden (vgl. S. 221). Wir betrachten daher jede untersuchte need-Kategorie für sich (z.B. nDom, vgl. Tab. IX). Wir gehen weiter - aufgrund obiger Voraussetzungen - davon aus, daß die Korr-Koeff. zwischen WP/real nicht derselben Grundgesamtheit angehören wie die Korr-Koeff. zwischen WP/ideal. Eine solche einzelne Form der Korrelation (z.B. die zwischen Dom WP / real-Dom) wird an 6 verschiedenen Stichproben (nämlich den 6 verschiedenartigen Extremgruppen, vgl. Tab. IX) durchgeführt. Doch liegt hier nicht der Fall vor, daß d i e s e l b e Hypothese an 6 Stichproben derselben Grundgesamtheit geprüft wird, denn wir erwarten für diese 6 Stichproben nicht dieselben Relationen; wir erwarten nicht für jede Stichprobe überhaupt eine sign. Relation. Wir gehen vielmehr theoretisch davon aus, daß diese

1) Die Signifikanz-Niveaus dieser Koeffizienten ergeben sich nur bei dem weniger strengen Maßstab der einseitigen Prüfung. Dieser Berechnung gehen zwar keine gerichteten Hypothesen voraus, doch läßt sich für alle Werte - post festum - zeigen, daß sie in der jeweils vorliegenden Richtung aufgrund der bereits dargestellten Ergebnisse erwartet werden können. Jedenfalls wird die Annahme entgegengesetzt gerichteter Korrelationen durch die bereits vorliegenden Ergebnisse ausgeschlossen (siehe dazu weiter unten).

Stichproben sich hinsichtlich des Verhältnisses von Index und Indiziertem unterscheiden; wir erwarten, daß sich die WP-scores bei diesen Stichproben nicht auf ein gemeinsames Kriterium beziehen, sondern getrennt validiert werden müssen.- Aus diesen theoretischen Erwägungen folgt, daß eine Zufallsbedingtheit der sign. Korr-Koeff., die auf die Äquivalenz der Stichproben zurückzuführen wäre, n i c h t gegeben ist. [1)]

Wir entnehmen den obigen Ergebnissen zunächst folgenden allgemeinen Hinweis : es läßt sich (im Rahmen unserer Versuchsanordnung) zeigen, daß - im Gegensatz zur gesamten Versuchsgruppe - bei einzelnen Untergruppen eine korrelative Beziehung zwischen den need-scores der WP und der Einschätzung derselben Variablen auf einer der Ebenen besteht.

Weiter ist folgender Zusammenhang bemerkenswert : Vpn, die sich durch extreme Selbsteinschätzung auf einer Ebene auszeichnen (z.B. geringste ideal-Aut-scores), werden durch ihre WP-Aut-scores danach differenziert, wie sie sich auf einer anderen Ebene (hier der real-Ebene) unterscheiden. Die einheitliche Selbsteinschätzung auf der einen Ebene (in diesem Falle besonders geringe ideal-Aut-scores) fungiert demnach als 'Trennkriterium', anhand dessen sich eine der gesuchten Untergruppen bilden läßt. Die Daten einer anderen Ebene sind es (hier die real-Ebene), mit welchen die Test-scores dieser Gruppe korrelieren, für welche die Test-scores also als 'Indikatoren' bewertet werden können.

Dieser Zusammenhang verdient eine nähere Analyse. Die Tatsache, daß Vpn mit besonders hoher bzw. besonders geringer Selbsteinschätzung auf einer Ebene zu Gruppen zusammengestellt werden, läßt sich dahingehend auffassen, daß in solch einer Gruppe ein Merkmal (nämlich die extreme Selbsteinschätzung) konstant gehalten wird. Wenn diese Vpn sich aber in ihrer Selbsteinschätzung auf einer anderen Ebene (also hinsichtlich eines anderen Merkmals) u n t e r s c h e i d e n , so kann dieser Umstand - wie gleich deutlich zu machen ist - dahingehend beschrieben werden, daß die Vpn sich

1) Es sei noch einmal in Erinnerung gebracht, daß wir hier nicht als Nullhypothese formulieren können, diese Stichproben gehörten derselben Grundgesamtheit an, um dann über die Gegenhypothese zu entscheiden, daß diese Stichproben verschiedenen Grundgesamtheiten zuzurechnen sind. Wir können nicht p r ü f e n , ob die Stichproben homolog sind; sondern es ist der experimentellen Anordnung zu entnehmen, ob sie in bezug auf die untersuchte Frage äquivalent sind oder nicht (vgl. dazu S.220).

hinsichtlich der D i f f e r e n z e n zwischen den betreffenden zwei Ebenen unterscheiden. In derselben Gruppe sind Vpn zusammengefaßt, die (dem obigen Beispiel nach) geringe ideal-Aut-scores und zugleich geringe real-Aut-scores haben und solche Vpn, die ebenfalls geringe ideal-Aut-scores, zugleich aber hohe real-Aut-scores aufweisen. Diese Vpn unterscheiden sich demnach nicht allein i.b. auf ihre real-Aut-scores, sondern sie unterscheiden sich insbesondere auch hinsichtlich der Differenzen zwischen ideal-Aut- und real-Aut-scores. Wenn der oben angegebenen Korrelation nach (Aut WP / real-Aut) die WP-Aut-scores um so höher sind, je größer die real-Aut-scores sind, so ist dem u.U. zu entnehmen, daß die WP-Aut-scores mit den Differenzen zwischen ideal- und real-Aut-scores variieren : denn diese Differenzen sind - so läßt sich vermuten - um so größer, je größer die real-Aut-scores ausfallen (bei einer Gruppe mit geringen ideal-Aut-scores).

Als weiteres Kriterium ist offenbar entscheidend, daß diese Vpn gleichartig durch geringe ideal-Aut-scores charakterisiert sind. Andernfalls nämlich wäre zu erwarten, daß dieselbe Relation auch für die Vpn mit den höchsten ideal-Aut-scores gilt, denn auch diese Vpn werden sich hinsichtlich der real-Ebene unterscheiden, so daß ebenfalls eine Gruppe mit unterschiedlichen Differenzen zwischen real//ideal (für nAut) zusammengestellt wäre.

Wir werden diese Frage weiter unten wieder aufgreifen. Zunächst wollen wir für die obigen sign. Korr-Koeff. eine kurze inhaltliche Interpretation geben. Hier ist es zusätzlich aufschlußreich, die Daten im einzelnen zu berücksichtigen, welche jeweils in eine Korrelation eingehen : für ein Verständnis dieser korrelativen Beziehungen ist es wichtig zu wissen, ob die Daten jener Ebene, mit welcher für eine Untergruppe die WP-scores korrelieren, durchgehend positiv (Zustimmung), oder negativ (Ablehnung) sind oder sich über die ganze Skala verteilen.[1] Die Korr-Koeff. beschreiben folgende Zusammenhänge[2] :

1) Es sei daran erinnert, daß die 9 Stufen der Q-sort-Skala als +4, +3 .....-3,-4 gekennzeichnet werden.
2) Die Korrelation für nAgg interpretieren wir an späterer Stelle im Zusammenhang mit den übrigen Werten zu dieser Variablen.

Vpn, die in ihrem Ich-Ideal nDom in extrem negativer Weise bewerten, haben dann auch einen geringen nDom-Ausdruck in der WP, wenn diese Variable gleichzeitig bedürfnismäßig irrelevant ist (negative Wunsch-Dom-scores); je mehr aber der negativen Bewertung im ideal ein starker Wunsch nach Dominanz gegenübersteht, um so stärker wird - sozusagen dem eigenen Wertmaßstab entgegen - nDom in der WP zum Ausdruck gebracht.

Vpn, die in vergleichbarer Weise nAut in ihren Ideal-Vorstellungen einen besonders geringen Platz zuweisen, zeigen nicht ebenso durchgängig auch einen geringen Motiv-Ausdruck in der WP : je mehr nämlich - wiederum dem Ich-Ideal entgegen - diese Vpn sich als autonom in ihrem Verhalten erleben (je größer die Differenz zwischen Verhalten und Verhaltensanspruch), um so stärker auch tritt diese Variable in der WP hervor.

Vpn, die ein besonders starkes Bedürfnis haben, bewundert zu werden (nExh), drücken diesen extremen Wunsch nicht direkt aus in der WP : der Motivausdruck wird vielmehr 'reguliert' durch das Ich-Ideal. Je mehr nämlich dieser Wunsch auch vom Ich-Ideal her zugelassen ist (positive ideal-scores), desto stärker tritt er in der WP in Erscheinung; je stärker dieses Bedürfnis aber durch das Ich-Ideal abgelehnt wird (negative ideal-Exh-scores), je weniger gelangt es zum Ausdruck.

Vpn, die sich als sehr wenig ängstlich erleben, bringen in ihren WP-Antworten umso weniger durch avoid charakterisierte Inhalte, je negativer sie diese Variable auch in ihrem Ich-Ideal bewerten. Je weniger extrem diese Ablehnung auf der Ideal-Ebene ist (es handelt sich insofern durchgehend um 'Ablehnung', als diese Vpn alle einen, wenn auch verschieden hohen, negativen ideal-avoid-score haben), desto häufiger treten Antworten auf, die mit avoid zu signieren sind.

Diese Ergebnisse lassen sich allgemein noch einmal folgendermaßen charakterisieren :

Bei verschiedenen Variablen läßt sich zeigen, daß für jeweils eine bestimmte Untergruppe ein korrelatives Verhält-

nis zwischen Test-Index und Selbsteinschätzung auf einer Ebene besteht. Eine solche Relation beinhaltet, daß die Test-scores die Vpn jeweils auf einer bestimmten 'Ebene' differenzieren, oder anders gewendet : daß für eine bestimmte Untergruppe die WP auf einer bestimmten 'Ebene' funktioniert. Es handelt sich allerdings um äußerst spezifische Gruppenmerkmale : die Gruppen werden auf Grund bestimmter Meßwerte für eine einzelne Variable (need-Kategorie) konstituiert; und die Normen, die sich auf Grund des korrelativen Zusammenhangs zwischen Test-Index und Kriterium bilden ließen, bezögen sich allein auf eben diese Variable (nDom etc.). Wir könnten demnach für eine bestimmte Gruppe wohl angeben, wie die WP etwa für die Variable nAut 'funktioniert', wüßten jedoch nicht, wie die nExh-scores bei dieser Gruppe auszuwerten wären.

Ein weiterer wichtiger Hinweis ist diesen Ergebnissen zu entnehmen : es lassen sich v e r s c h i e d e n e Ebenen in einen korrelativen Zusammenhang mit den Test-Indizes bringen. Je nach Variable und Untergruppe sind die Test-scores als Ausdruck der Ich-Ideal-Vorstellungen zu interpretieren, oder als Hinweis auf unterschiedliche Einschätzungen auf der Wunsch-Ebene zu betrachten oder schließlich als Indizes für unterschiedliches Verhalten (wie es 'erlebt' wird) zu werten. Es zeigt sich also auch hier, daß mit dem Konzept verschiedener 'Ebenen' ein entscheidender Gesichtspunkt eingeführt ist; die Erwartung verschiedener 'Funktionsebenen' erweist sich - für dieses Material - als zutreffend.
Bevor wir diese Ergebnisse mit den in Abschnitt c) erörterten in einen Zusammenhang bringen können, ist eine Frage wieder aufzunehmen, die wir oben zunächst zurückgestellt hatten. Den Korr-Koeff. wurde der Hinweis entnommen, daß für den Grad des Motivausdrucks in der WP der Grad der Differenz zwischen 2 Ebenen ausschlaggebend sei (vgl.S.247). Dabei blieb offen, ob der Grad der Differenz allein als Kriterium zu werten oder ob zusätzlich ausschlaggebend sei, daß die Vpn sich durch eine extreme Selbsteinschätzung auf einer der Ebenen auszeichnen.

Zwei Fragen sind in diesem Zusammenhang noch zu klären. Wir haben die indirekt abgeleitete Behauptung, daß die Differenzen ausschlaggebend seien, d i r e k t , nämlich durch

die Diff-scores selbst, zu prüfen; und wir haben zu untersuchen, ob die Rolle der Differenzen nur für die in den obigen Korr-Koeff. herangezogenen E x t r e m gruppen gilt oder für die ganze Versuchsgruppe. Der Grad der Selbsteinschätzung wäre dann, wie vermutet, als 'Trennkriterium' zu bewerten, wenn der beschriebene korrelative Zusammenhang tatsächlich nur für die betreffende Extremgruppe nachweisbar ist.

Die erste der beiden Fragen ist folgendermaßen zu untersuchen :

2. Läßt sich für dieselben Vpn, für welche sich ein korrelativer Zusammenhang zwischen WP und Q_s-sort-Ebene nachweisen ließ, eine Korrelation nachweisen zwischen WP und den als ausschlaggebend behaupteten Differenzen zwischen 2 Ebenen?

Das Vorgehen sei beispielhaft an einer Variablen erläutert. Nach Tab. IX korrelieren die WP-Dom-scores mit den Wunsch-Dom-scores bei jenen Vpn, die sich durch geringe ideal-Dom-scores auszeichnen. Diese Relation wurde dahingehend interpretiert, daß die WP-Dom-scores um so größer sind, je mehr Wunsch-Dom > ideal-Dom, d.h. je größer die Differenz zwischen Wunsch-Dom//ideal-Dom. Wir berechnen nun für dieselben Vpn die Korrelation zwischen Dom WP und den Diff-scores Wunsch-Dom//ideal-Dom. Entsprechend gehen wir für die Variablen nAut, nExh, nAgg und avoid vor, indem für die in den Korr-Koeff. indirekt angezeigten Differenzen die entsprechenden Diff-scores herangezogen werden.

Es ergeben sich folgende sign. Korrelationen :

Vpn mit den geringsten ideal-Dom-scores :

Dom WP / Wunsch-Dom//ideal-Dom : r_s=+0,654, $p < 0.05$ [1)]

Vpn mit den höchsten Wunsch-Exh-scores :

Exh WP / Wunsch Exh//ideal-Exh : r_s=-0,73, $p < 0.05$

Für die übrigen Variablen ergeben sich zwar Korrelationen in der erwarteten Richtung, doch lassen sie sich nicht auf dem 5%-Niveau sichern. Die obigen Korr-Koeff. beschreiben dieselben Relationen, die sich oben (s.S.245) ergaben :

Für die Vpn mit den geringsten ideal-Dom-scores gilt, daß sie um so höhere nDom-scores in der WP aufweisen, je größer die Differenz ist zwischen Wunsch-Dom und ideal-Dom, d.h. je höher die Wunsch-Dom-scores gegenüber den ideal-Dom-scores sind.

---

1) Da gerichtete Hypothesen vorliegen, genügte hier eine einseitige Signifikanzprüfung.

Für die Vpn mit den höchsten Wunsch-Exh-scores gilt, daß sie um so geringere nExh-scores in der WP erhalten, je größer die Differenz ist zwischen Wunsch-Exh und ideal-Exh, d.h. je geringer die ideal-Exh-scores gegenüber den Wunsch-Exh-scores sind.

Die Interpretation, daß in den auf S.245 angegebenen Korr-Koeff. eine Relation zwischen WP-scores einerseits und Differenzen andererseits angezeigt würde, hält demnach nur für 2 Variablen einer näheren Prüfung stand. Wie sind die Korr-Koeff. für die Variablen nAgg, nAut, und avoid zu interpretieren?

Der Unterschied in den Ergebnissen läßt sich folgendermaßen erklären. Wenn Extremgruppen gebildet werden, z.B. die Vpn mit den geringsten ideal-Aut-scores, so ist auch in dieser Gruppe noch eine gewisse Streuung der ideal-Aut-scores gegeben : Nur für den Fall aber, daß diese ideal-Aut-scores (annähernd) identisch sind, können – in diesem Beispiel – inter-individuelle Unterschiede in den real-Aut-scores gleichgesetzt werden mit inter-individuellen Unterschieden in der Differenz zwischen den ideal-scores und den real-scores. Andernfalls sind hohe real-Aut-scores (in diesem Beispiel) nicht notwendig auch mit einer hohen Differenz zwischen real- und ideal-Aut-scores verknüpft. Die Interpretation, daß bei diesen Extremgruppen inter-individuelle Unterschiede auf der einen Ebene inter-individuellen Unterschieden in der Differenz zwischen 2 Ebenen entsprächen, ist also genaugenommen nur gerechtfertigt für den Fall, daß die Vpn auf der Ebene, die als Kriterium fungiert, tatsächlich gleiche scores haben.

Es werden in diesen Korr-Koeff. also zwei verschiedenartige Zusammenhänge beschrieben :

Im einen Falle besteht eine Korrelation zwischen Motivausdruck in der WP und Grad der Differenz zwischen 2 Ebenen (für nDom und nExh).

Im anderen Falle korrelieren für die Vpn mit extremer Selbsteinschätzung auf einer Ebene die WP-scores mit den scores einer anderen Ebene (nAgg, nAut, avoid).

Diese Verschiedenartigkeit ist auf folgenden Sachverhalt zurückzuführen.

Im ersteren Fall besteht zusätzlich eine Korrelation zwischen dem Grad der Differenz zwischen 2 Ebenen und den scores einer der beiden Ebenen (diese Korrelation läßt sich für beide Variablen auf dem 1%-Niveau sichern). Für das Beispiel der Variable nDom heißt das : bei den Vpn mit den geringsten ideal-Dom-scores bestehen folgende Korrelationen :

Dom WP / Wunsch-Dom (vgl.S.245)
Dom WP / Wunsch-Dom//ideal-Dom (vgl.S.250)
Wunsch-Dom / Wunsch-Dom//ideal-Dom (r_s=0,906)

Für diese Variablen ist demnach der oben beschriebene Sonderfall gegeben, daß die Differenzen um so größer ausfallen, je größer die scores einer der beiden Ebenen sind.

Im zweiten Fall ist eine solche Korrelation zwischen den Diff-scores zweier Ebenen und den scores einer der beiden Ebenen nicht gegeben und entsprechend nicht eine Korrelation zwischen WP-scores und Diff-scores.

Aus dieser Gegenüberstellung ist der Hinweis zu entnehmen, daß (in u n s e r e m Material) der Korrelation zwischen WP-scores und den scores einer Ebene bei den Vpn, die sich durch eine extreme Selbsteinschätzung auf einer der 3 Ebenen auszeichnen, die allgemeinere Bedeutung zukommt : nur in 2 Fällen (nDom, nExh) entspricht dieser Korrelation zugleich eine Korrelation zwischen WP-scores und dem Grad der Differenz zwischen 2 Ebenen.

Es könnte sich allerdings auf der Grundlage eines umfangreicheren Datenmaterials die hier als Sonderfall beschriebene Relation als die allgemeinere erweisen : nämlich für den Fall, daß in einer breiter angelegten Untersuchung eine größere Gruppe von Vpn mit extremer Selbsteinschätzung zusammengestellt werden könnte, für welche zugleich dieses Merkmal (die extreme Selbsteinschätzung) nur unerheblich streut. Hier könnte sich ergeben, daß zwar auch - wie bei unserem Datenmaterial - die Daten e i n e r Ebene mit jenen der WP korrelieren, daß jedoch die Korrelation zwischen Diff-scores und WP-scores eine höhere, vielleicht statistisch besser gesicherte wäre.

Die Frage, welcher der beiden beschriebenen Relationen die größere Rolle zukommt, kann im Rahmen dieser Untersuchung nicht entschieden werden. Doch ist mit unseren Daten jedenfalls dieses Problem gestellt. Diagnostisch ist es von einiger Bedeutung zu wissen, ob unterschiedliche Testreaktionen eher Unterschiede z.B. hinsichtlich des Ich-Ideals indizieren, oder ob sie als Hinweis für einen unterschiedlichen Grad der 'Spannung' zwischen 2 Ebenen zu werten sind.

Für unser Datenmaterial bleibt nun weiter zu klären, ob die korrelative Beziehung zwischen WP- und Diff-scores nur auf die Extremgruppen beschränkt ist, oder ob etwa in dieser Korrelation ein allgemeinerer Zusammenhang angezeigt ist, der für die gesamte Versuchsgruppe gilt. Somit ergibt sich die Frage :

3. Besteht die korrelative Relation zwischen WP-scores und Diff-scores, die sich für einzelne Untergruppen nachweisen ließ, gleichermaßen für alle Vpn?

   Diese Frage ist insbesondere für die Variablen nExh und nDom zu klären; wir prüfen sie aber auch für jene Variablen, bei welchen diese Rolle der Ebenen-Differenzen nicht hat nachgewiesen werden können (nAgg,nAut,avoid). Wir gehen dabei entsprechend vor, wie es zur Frage 2) beschrieben wurde. Es ist die Nullhypothese zu prüfen, daß eine derartige Korrelation für die gesamte Versuchsgruppe nicht besteht.

Eine sign. Korrelation zwischen WP-scores und Diff-scores ergibt sich allein für die Variable nDom : r_s=0,526 [1]. Dieselbe Korrelation haben wir schon in einem anderen Zusammenhang diskutiert und inhaltlich interpretiert (vgl.S.239).

Auf der Grundlage unseres Materials läßt sich nicht entscheiden, weshalb es gerade die Variable nDom ist, für welche sich ein korrelativer Zusammenhang zwischen WP-Dom-scores und Diff-scores für die gesamte Versuchsgruppe nachweisen läßt. Auch können wir im Rahmen dieses Experiments nicht untersuchen, warum ein vergleichbarer korrelativer Zusammenhang für andere Variablen nicht gegeben ist.

Aus der Tatsache jedoch, daß diese korrelative Beziehung zwischen Diff- und WP-scores - mit Ausnahme einer Variablen - für die gesamte Versuchsgruppe n i c h t nachweisbar ist,

1) Dieser Wert ist bei einseitiger Prüfung auf dem 1%-Niveau gesichert.

läßt sich die Vermutung herleiten, daß es nicht die 'absolute' Größe der Differenz zweier Ebenen allein ist, welche mit den WP-Reaktionen in Verbindung zu bringen ist, sondern daß ein weiteres Merkmal berücksichtigt werden muß. In zwei Fällen hat sich eine solche Korrelation für solche Vpn ergeben, welche durch das einheitliche Merkmal einer extremen Selbsteinschätzung charakterisiert sind.

Diese Annahme, daß eine Relation zwischen WP- und Diff-scores nur dann zu erwarten sei, wenn ein zusätzliches Merkmal erfüllt wäre, ist abzuwägen gegen eine andere bereits aufgewiesene Konsequenz des Datenmaterials. Den Tabellen VI und VII zufolge liefern die Diff-scores, verwendet man sie als dichotomisiertes Merkmal, eine statistisch gut gesicherte inter-individuelle Differenzierung. Für die Variablen nDom, nNur und avoid ließ sich diese Relation kreuzweise sichern : es ergab sich dieselbe Relation, ob man nun Stichproben auf Grund der WP-scores oder auf Grund der Diff-scores bildete. Demnach spielen hohe bzw. geringe Differenzen als s o l c h e eine Rolle, u n a b h ä n g i g von einer zusätzlichen Bedingung.

Doch ist hier einzuwenden, daß diese für die Extremgruppen kreuzweise gesicherte Relation (in unserem Material) n i c h t als Ausdruck einer korrelativen Beziehung gewertet werden kann : denn eben diese Korrelation ließ sich für die gesamte Gruppe nur bei nDom nachweisen. Demnach leisten die Diff-scores wohl eine 'grobe' inter-individuelle Differenzierung : sofern aber überhaupt feinere inter-individuelle Unterschiede durch Diff-scores widergespiegelt werden, ist damit zu rechnen, daß solche nur für U n t e r -gruppen nachweisbar sind.

Damit kehren wir zu der obigen Überlegung zurück. Unserem Material nach ist es wahrscheinlicher, daß nicht die Differenzen als solche auf die WP-Reaktionen zu beziehen sind, sondern erst im Zusammenhang eines weiteren Merkmals resp. Trennkriteriums. Untergruppen, die auf diese Weise zu bilden sind, sind dadurch charakterisiert, daß die Vpn hinsichtlich eines Merkmals (des 'Trennkriteriums') homogen sind und sich hinsichtlich jenes Merkmals unterscheiden, für welches die Testreaktionen als Indikatoren zu bewerten sind.

Diese Überlegung führt zu einem weiteren Problemkomplex, der jedoch hier ausgeklammert bleiben soll. Es ist zu beachten, daß der 'absolute' Vergleich von Differenzen eine - u.U. - entscheidende Information unterschlägt : eine geringe Differenz kann Ausdruck zweier etwa gleich hoher Einschätzungen auf der positiven wie auf der negativen Seite der Skala sein. Vpn, die sich hinsichtlich einer Variablen in real und Wunsch hoch positiv einschätzen, unterscheiden sich - in den Diff-scores - nicht von den Vpn, welche diese Variable auf beiden Ebenen etwa gleich stark ablehnen.

Dieser Unterschied wird berücksichtigt, wenn eine der beiden Ebenen, zwischen denen die Differenz gebildet wurde, 'konstant gehalten' wird, d.h. wenn nur Vpn zu einer Gruppe zusammengestellt werden, die hinsichtlich einer der beiden Ebenen übereinstimmen, z.B. darin, daß sie diese Ebene übereinstimmend besonders hoch einschätzen.

Unberücksichtigt bleibt dieser Unterschied in dem Falle, daß für diese Diff-scores ein Extremgruppenvergleich durchgeführt wird (vgl. Tab. VI) : die (Extrem-)Gruppe etwa von Vpn mit besonders geringen Differenzen ist in der oben beschriebenen Weise unterschiedlich zusammengesetzt. Hier aber spielt es offenbar keine Rolle, welchem Skalenbereich die Differenzen zugeordnet sind : denn für diese Extremgruppen läßt sich eine systematische Relation zwischen WP-scores und Diff-scores nachweisen.

Es ist demnach ungeklärt, wann die Differenzen 'als solche' von Bedeutung sind und wann die Differenz in Relation zu dem Skalenbereich, dem die zwei verglichenen Einschätzungen zugehören, zu betrachten sei. Diese Frage einer Interaktion zwischen 'Differenz' und 'Skalenbereich' wäre daher im einzelnen für die gesamte Skala zu analysieren : sie kann indessen im Rahmen dieser Arbeit nicht mit berücksichtigt werden.

Noch eine weitere Frage läßt sich an die Diff-scores knüpfen. Dem Auswertungsgang, der zu Tabelle IX führte (vgl. S. 243ff), formal entsprechend, wäre zu prüfen, ob die Diff-scores, ähnlich den Selbsteinschätzungen auf einer Ebene, als 'Trennkriterium' fungieren.

Vermittels Extremgruppenvergleich hatten wir (für einzelne Variablen) festgestellt, daß Vpn mit hohen bzw. geringen Diff-scores sign. verschieden auf die WP reagieren, d.h. sich in der Höhe ihrer need-scores (i.b. auf die betreffende Variable) unterscheiden. 'Auf welcher Ebene' aber sind diese need-scores zu interpretieren : lassen sie sich nur, wie es dem Extremgruppenvergleich zu entnehmen ist, auf die Alternative 'hohe bzw. geringe Differenzen zwischen 2 Ebenen' beziehen; oder läßt sich eine feinere inter-individuelle Differenzierung nachweisen, d.h. eine Differenzierung auf einer bestimmten Ebene? Gilt etwa für Vpn mit besonders hohen Differenzen zwischen real//Wunsch z.B. für nAgg, daß die WP-scores diese Vpn danach differenzieren, wie sie ihr V e r h a l t e n einschätzen oder welcherart Aggressions - W ü n s c h e sie haben? Es ist also folgende Frage zu prüfen :

4. Läßt sich für Vpn mit besonders hohen bzw. besonders geringen Differenzen zwischen 2 Ebenen eine Korrelation nachweisen zwischen WP-scores und den scores einer der beiden Ebenen für die betreffende Variable?

Wir sind auch dieser Frage nicht im einzelnen nachgegangen, sondern haben sie nur probeweise an einigen Variablen untersucht, ohne eine derartige Relation festzustellen. Zwar ließe nur eine vollständige Analyse des Datenmaterials eine Entscheidung darüber zu, ob auch den Diff-scores die Rolle von Trennkriterien zukommt; doch sei eine theoretische Überlegung an die Stelle dieses - ziemlich aufwendigen - Auswertungsschritts gesetzt.

Auch mit dieser Fragestellung nämlich ist das oben beschriebene Problem verknüpft : das Problem der Interaktion von 'Differenz' und 'Skalenbereich'. Eine Gruppe mit besonders hohen Differenzen ist - anders als eine Gruppe mit extremer Selbsteinschätzung auf e i n e r Ebene - nicht durch ein einheitliches Merkmal charakterisiert. Besonders hohe Differenzen kennzeichnen sowohl Vpn mit hoher Selbsteinschätzung auf Ebene 'a' und mittleren scores auf Ebene 'b', wie Vpn mit mittlerer Einschätzung auf Ebene 'a' und besonders starker Ablehnung auf Ebene 'b'. Als Beispiel gefaßt bedeutet das : dieselbe 'absolute' Differenz charakterisiert sowohl Vpn,

die sich hinsichtlich nAgg im real extrem hoch und das ideal im mittleren Bereich einschätzen, wie Vpn, die ihr Verhalten als 'durchschnittlich' angeben, jedoch Aggressivität auf der Ideal-Ebene besonders scharf ablehnen. Es ist zweifelhaft (wenn auch nur empirisch entscheidbar), daß diese Vpn hinsichtlich ihrer WP-Reaktion derselben Grundgesamtheit zuzuordnen sind; d.h. dieser Argumentation nach ist es unwahrscheinlich, daß hohe bzw. geringe Differenzen eine Gruppe gleichartiger Reaktionen auf die WP konstituieren, also als Trennkriterien fungieren.

Hier mag jedoch eingewendet werden, daß die für a l l e Vpn gesicherte Korrelation zwischen Diff-scores für nDom und den WP-Dom-scores (vgl. S. 239) das Gegenteil beweist. Denn diese Korrelation gilt für die Diff-scores ihrem 'absoluten' Betrag nach : der jeweils zugehörige Skalenbereich bleibt unberücksichtigt; kein Trennkriterium wurde hinzugezogen, kein Merkmal 'konstant' gehalten (vgl.S.255). Diese Korrelation sei - ihrer Unvereinbarkeit mit der obigen Annahme wegen - noch einmal näher analysiert.

Wir gehen - wie mehrfach beschrieben - in dieser Annahme davon aus, daß hohe bzw. geringe Diff-scores verschiedenen Skalenbereichen zugeordnet sind, so daß in dieser Korrelation zwischen Diff-scores und WP-Dom-scores sowohl hohe Differenzen, die dem positiven Skalenbereich, wie hohe Differenzen, die dem negativen Skalenbereich zugehören (vgl. das obige Beispiel), mit hohen WP-Dom-scores verknüpft wären.

Es ist daher zu prüfen, ob diese Annahme auf die vorliegende Korrelation (Dom WP / Wunsch-Dom//ideal-Dom) überhaupt zutrifft oder ob die Höhe der Differenz in diesem Falle n i c h t unabhängig ist vom zugehörigen Skalenbereich; d.h. ob gleich hohe Differenzen hier - unserer Erwartung entgegen - demselben Skalenbereich zugehören. Für diesen Fall wäre eine korrelative Beziehung zwischen Differenzen und Skalenbereich zu erwarten. Wir prüfen daher die folgende Frage :

5. Besteht eine korrelative Beziehung zwischen den Diff-scores Wunsch-Dom//ideal-Dom und den scores einer der beiden Ebenen?

Es ist also einmal die Korrelation festzustellen zwischen Diff-scores und W u n s c h-Dom-scores und jene zwischen Diff-scores und i d e a l-Dom-scores. Es ergibt sich eine hoch sign. Korrelation zwischen den Diff-scores und den Wunsch-Dom-scores :

Wunsch-Dom / Wunsch-Dom//ideal-Dom : $r_s = 0{,}684$
($t=5{,}3$ $df=32$ $p< 0.0005$)

Diese Korrelation beschreibt den Sonderfall, daß der Grad der Differenz zwischen 2 Ebenen mit dem Grad der Ausprägung des Merkmals auf einer der beiden Ebenen variiert. Im obigen Fall der Korrelation zwischen Diff-scores und WP-Dom-scores besteht die korrelative Beziehung demnach n i c h t unabhängig vom Skalenbereich : hier ist nicht der Fall gegeben, daß gleich hohe Differenzen verschiedener Skalenbereiche mit denselben WP-scores verknüpft sind; denn gleich hohe Differenzen gehören hier dem gleichen Skalenbereich an.

Die Korrelation zwischen Diff-scores und WP-Dom-scores (für alle Vpn) widerspricht somit nicht der Annahme, daß die Diff-scores in Abhängigkeit vom Skalenbereich zu analysieren sind. Sie widerlegt nicht - wie zunächst als Einwand formuliert worden war - die Vermutung, daß Vpn mit einer hohen Differenz im positiven Skalenbereich und solche mit einer hohen Differenz im negativen Skalenbereich n i c h t (i.b. auf die WP-Reaktion) derselben Grundgesamtheit zuzurechnen sind.

Gleichwohl lassen sich dieser Diskussion keine positiven Hinweise auf die Art der Interaktion von 'Differenz' und 'Skalenbereich' entnehmen. Wir sehen uns jedoch in der Annahme bekräftigt, daß die Differenzen keine Trennkriterien für unterschiedliche Relationen zwischen Test-Index und Kriterium liefern, sondern daß Untergruppen eher durch das Merkmal einer einheitlichen Selbsteinschätzung auf e i n e r Ebene konstituiert werden.

Die intra-individuellen Differenzen zwischen 2 Ebenen spielen eher eine Rolle als (inter-individuell) differenzierendes Kriterium, denn als vereinheitlichendes Merkmal. Für die Diff-scores ließ sich, verwendete man sie als dichotomisiertes Merkmal, der Nachweis einer inter-individuellen

Differenzierung am besten sichern (Vergleich der Tabellen VI und VII; s.S. 238). Bemerkenswert ist ferner, daß wohl die Korrelation Dom WP / Wunsch-Dom//ideal-Dom für die gesamte Versuchsgruppe gilt, nicht jedoch eine Korrelation zwischen den WP-scores und den scores nur e i n e r Ebene; auch hier läßt sich ein Zusammenhang zwischen WP-scores und Kriterium gerade für den Fall sichern, daß man die Diff-scores verwendet.

Zur Rolle des Ebenen-Gefüges, betrachtet für alle Variablen gleichzeitig.

Die Relevanz der Differenzen zwischen 2 Ebenen für die Analyse von Testreaktionen haben wir bisher nur untersucht für einzelne Variablen : einzelne Gruppierungen von Vpn wurden auf der Grundlage der Differenzen für e i n e Variable zusammengestellt. Die Q_s-sort-Methode ist (unseres Wissens) für eine derartige Auswertung bisher nicht herangezogen worden. In der vorliegenden Literatur dienen vielmehr jene Differenzen als Ausgangsdaten, welche (entweder als Diff-scores oder als Korr-Koeff. berechnet) das Verhältnis der Einschätzungen zweier Ebenen insgesamt gesehen wiedergeben. Wir haben oben schon einmal darauf Bezug genommen, daß diesen 'Ges.Korr.Koeff.' von einer Vielzahl von Autoren eine diagnostische Bedeutung zugesprochen wird (s.S. 240). Es wird zumeist angenommen, daß eine hohe Übereinstimmung zwischen 2 Ebenen (in diesen Untersuchungen werden nur real- und ideal-self erhoben) 'gute Angepaßtheit' indiziert, während eine geringe Übereinstimmung auf 'schlechte Angepaßtheit' verweist.

Wir wollen nun hier 'erkunden', ob dieser Grad der Übereinstimmung zwischen 2 Ebenen in Zusammenhang zu bringen ist zur Testreaktion, ob insbesondere in diesen Ges.Korr. Koeff. Bedingungen für unterschiedliche Testreaktionen beschrieben werden. Führt etwa eine hohe Übereinstimmung zwischen 2 Ebenen dazu, daß die WP dem 'Seinsbestand' (Wilde) nach beantwortet wird, daß also das 'Verhalten' in den Testreaktionen widergespiegelt wird; ist eine hohe Diskrepanz zwischen 2 Ebenen eine Bedingung dafür, daß die (nicht erfüllten) Bedürfnisse 'projiziert' werden oder

daß man seine Antworten an den (nicht-realisierten) Ich-Ideal-Normen ausrichtet? Auch hier lassen sich keine zureichend begründeten Hypothesen aufstellen, sondern eher eine Vielfalt - einander ausschließender - Möglichkeiten formulieren. Es kann also nur 'probeweise' untersucht werden, ob den Ges.Korr.Koeff. die Rolle von Trennkriterien zukommt : es wird allgemein gefragt, ob durch dieses Merkmal Gruppen konstituiert werden, für welche die Test-scores inter-individuelle Unterschiede auf einer bestimmten Ebene (ausgedrückt durch eine korrelative Beziehung) widerspiegeln.

Die Gruppen, welche auf der Grundlage dieser Ges.Korr.Koeff. gebildet werden können, unterscheiden sich von jenen, welche in Orientierung an den Diff-scores zusammengestellt wurden, in einem wesentlichen Punkt : gleich hohe (bzw.geringe) Ges.Korr.Koeff. stellen - im Gegensatz zu ihrem absoluten Wert nach gleich großen Diff-scores (vgl. S. 255f) - ein einheitliches Merkmal dar; d.h. die Gruppen, die anhand dieses Merkmals zusammengestellt werden, sind - i.b. auf dieses Merkmal - in der Tat homogen.

Folgende Frage ist hier - analog zur Frage 1) dieses Abschnitts (vgl. S. 243) - zu stellen :

6. Bilden Vpn mit gleichartigen Ges.Korr.Koeff. zwischen 2 Ebenen eine Gruppe, für welche sich eine Korrelation zwischen Test-Index und Kriterium nachweisen läßt?

   Es sind hier die Daten von 3 verschiedenen Ges.Korr.Koeff. heranzuziehen (nämlich der Ges.Korr.Koeff. zwischen real/ideal, real/Wunsch, ideal/Wunsch). Auch hier werden Extremgruppen gebildet, so daß sich insgesamt 6 Gruppierungen ergeben. Für diese Gruppen wird jeweils bei allen Variablen geprüft, ob eine Korrelation zwischen Test-Index einerseits und Selbsteinschätzung andererseits nachweisbar ist. Ähnlich dem unter Frage 5) des vorhergehenden Abschnitts (s.S. 240f) beschriebenen Auswertungsschritt werden demnach auch hier für eine einzige Gruppe eine Vielzahl von Korrelationen durchgeführt (s.u.).

Für die nachfolgend aufgeführte Tabelle ist ein besonderer Hinweis nötig. Obwohl bei dieser Auswertung der strengere Maßstab der zweiseitigen Prüfung anzulegen ist (da gerichtete Hypothesen fehlen), haben wir die Werte, die bei einseitiger Prüfung noch auf dem 5%-Niveau sign. wären, mit angeführt, jedoch durch Klammern ( ) von den übrigen abgehoben. Das

Signifikanz-Niveau der zweiseitig geprüften Korr.Koeff. ist an folgenden Werten abzulesen :

$r_s \geq 0{,}66$ $\quad p < 0.05$
$r_s \geq 0{,}746$ $\quad p < 0.02$
$r_s \geq 0{,}80$ $\quad p < 0.01$

Für diese - in Tabelle X (s.S.262) angeführten - Korrelationen ist zunächst wieder zu prüfen, ob es sich um zufallsbedingt signifikante Abweichungen von der Nullhypothese handelt. Diese Frage kann hier ebenso wenig präzise beantwortet werden wie für die Werte der Tabelle VIII (s.S.241). Denn hier werden - ebenso wie dort - für ein und dieselbe Gruppe von Vpn sehr viel mehr Signifikanztests durchgeführt als bei den übrigen Tabellen. Für die Annahme, daß es sich dabei n i c h t um äquivalente Stichproben handelt, gibt es auch hier keine theoretisch hinlänglich stichhaltigen Beweise. Daß wir diese Frage trotzdem eingehender behandeln, hat den folgenden Grund : in der Gegenüberstellung dieses Auswertungsschritts mit den übrigen kann die Eigenart der bisher diskutierten Ergebnisse noch einmal schärfer umrissen und kann abgewogen werden, für welche Fragestellungen das Material am ehesten verwertet werden kann.

Kennzeichnend ist für diesen Auswertungsschritt vor allem, daß es zweifelhaft ist, ob die Daten verschiedener Variablen - i.b. auf das verwendete Kriterium - unterschiedlichen Grundgesamtheiten zugerechnet werden können (vgl. S.221). So gibt es z.B. keinen Ansatzpunkt für die Annahme, daß Vpn, die sich durch eine hohe Übereinstimmung von Wunsch- und ideal-Ebene auszeichnen (Vpn mit hohen Ges.Korr.Koeff. Wunsch / ideal), hinsichtlich der Variablen nAgg anders auf die WP reagieren als hinsichtlich nNur. Oder es liegen entsprechend keine empirischen Anhaltspunkte vor für die Annahme, daß Vpn, bei denen Wunsch- und ideal-Ebene stark auseinanderfallen, hinsichtlich der Variablen nDom im Sinne der ideal-Ebene auf die WP reagierten und in ihren nAut-Antworten ihre 'Wünsche' zum Ausdruck brächten.

Soweit wir jedoch - bei den übrigen Auswertungsschritten - Kriterien verwendeten, die sich auf die einzelnen Variablen

Tab. X : Extreme Ges.Korr.Koeff. als Trennkriterien

(Korrelation zwischen WP-scores und Q_s-sort-scores)

Niedrigste Ges.Korr.Koeff.

Variable	real / ideal	real / Wunsch	ideal / Wunsch
-Agg		WP/Wunsch r_s=+0,673	WP/ideal r_s=-0,76
+Agg		WP/ideal r_s=-0,66	
-Agg/+Agg		WP/ideal r_s=+0,73	WP/ideal r_s=+0,654
Dom			
Nur			
Aff		WP/real (r_s=+0,603)	
Aut		WP/Wunsch (r_s=+0,573)	
Exh			
avoid			

Höchste Ges.Korr.Koeff.

Variable	real / ideal	real / Wunsch	ideal / Wunsch
-Agg			
+Agg			WP/real r_s=-0,74
-Agg/+Agg	WP/ideal (r_s=+0,576)		
Dom			
Nur	WP/ideal r_s=+0,882		WP/ideal r_s=+0,754
Aff			
Aut			
Exh			WP/ideal (r_s=+0,633)
avoid	WP/ideal (r_s=+0,63)	WP/ideal (r_s= 0,60)	WP/ideal (r_s=+0,594)

getrennt bezogen, war das soeben beschriebene Problem nicht gegeben : die Annahme, daß für verschiedene Variablen verschiedenartige Relationen zwischen Index und Indiziertem zu erwarten sind, ist - im Falle solcher an den einzelnen Variablen ausgerichteter Kriterien - durch die in Kap. II referierten experimentellen Befunde theoretisch hinlänglich gestützt. So läßt sich aus den vorliegenden Experimenten die Erwartung ableiten, daß ein positives Selbstbild oder ein stark ausgeprägter 'Wunsch' i.b. auf eine sozial n e g a t i v bewertete Variable sich anders auf die Testreaktion auswirken als ein positives Selbstbild bzw. ein starker 'Wunsch' hinsichtlich einer sozial p o s i t i v bewerteten Variablen : hier könnte man etwa erwarten, daß die Testreaktionen im ersten, im Gegensatz zum zweiten Falle durch Verteidigungsverhalten gekennzeichnet sind. Ebenso läßt sich vermuten, daß inter-individuelle Unterschiede hinsichtlich der Differenz zwischen real//ideal in den WP-Reaktionen verschiedenartig widergespiegelt werden, je nachdem ob bei der betrachteten Variablen das ideal durchgehend hoch (z.B.nAut) oder von allen Vpn negativ (avoid) eingeschätzt wird.

Wenn aber ein einheitliches Merkmal (Ges.Korr.Koeff.) auf verschiedene need-Kategorien angewendet wird, lassen sich keine Annahmen darüber begründen, in welcher Hinsicht sich das Verhältnis von Index (WP) und Kriterium (Selbsteinschätzung) bei diesen needs unterscheiden soll.

Für jede der 6 Untergruppen werden die scores von 8 Variablen mit den Daten der real-, der ideal- und der Wunsch-Ebene korreliert. Der bisherigen Argumentation folgend können wir auch hier davon ausgehen, daß die Korrelationen mit den verschiedenen 'Ebenen' nicht derselben Grundgesamtheit angehören. Ob aber die 8 Korrelationen zwischen WP-scores und e i n e r Ebene (bei 8 Variablen) als homolog anzusehen sind oder nicht, können wir nicht entscheiden, weil es theoretisch ungeklärt ist, ob sich das einheitliche Merkmal (Ges.Korr.Koeff.) gleichartig auf die Testreaktionen auswirkt oder nicht. Schließlich kann nicht entschieden werden, ob die 6 verschiedenen Untergruppen i.b. auf die Testreaktionen verschiedenen Grundgesamtheiten zugerechnet werden können; denn es gibt keine theoretischen Anhaltspunkte dafür, ob und wie sich das Verhältnis zweier 'Ebenen insgesamt betrachtet' auf die Testreaktionen auswirken könne, analysiert man sie für die einzelnen Variablen getrennt.

Doch handelt es sich andererseits bei dieser Auswertung nicht um eine - an abhängigen Stichproben - wiederholte Prüfung d e r s e l b e n Hypothese : denn eine für alle Variablen gleichartig gerichtete Hypothese liegt ebenfalls nicht vor. Wir 'erkunden' nur, o b dem betrachteten Merkmal, den Ges.Korr.Koeff., innerhalb unseres Materials die Rolle von Trennkriterien zukommt. Solche 'blind analysis' aber erweist sich hier,wie im Falle der Tabelle VIII, wieder als unfruchtbar : es läßt sich - so weit wir sehen - nicht entscheiden, welche Stichhaltigkeit den Ergebnissen zuzusprechen ist.

Mit allem nötigen Vorbehalt wollen wir der Tabelle X nur einige formale Hinweise entnehmen; eine inhaltliche Interpretation erscheint uns nicht gerechtfertigt. Zunächst einmal ist festzustellen, daß die Annahme, das 'Gefüge zweier Ebenen insgesamt gesehen' lasse sich in Zusammenhang bringen mit den Testreaktionen, nicht als widerlegt gelten muß : es ist zu beachten, daß einzelne Korrelationen sich auf dem 1%- bzw. 2%-Niveau sichern ließen. Gleichwohl können wir nicht positiv folgern, daß unter anderen experimentellen Bedingungen mit einem derartigen Zusammenhang gerechnet werden muß.

Im einzelnen ist auf folgende Werte der Tabelle X zu verweisen :

Für den Quotienten $^{-Agg}/_{+Agg}$ ergibt sich in 3 Fällen eine Korrelation mit den ideal-Agg-scores. Dieselbe Korrelation wird insgesamt an 6 unabhängigen Stichproben durchgeführt : die Zufallswahrscheinlichkeit von 3 sign. Korrelationen betrüge hier - bei Voraussetzung der Äquivalenz der Stichproben - $p < 0.0022$ (Wilkinson 1951). Bemerkenswert ist dabei, daß dieselbe Relation für Gruppen mit geringen Ges.Korr.Koeff. wie für eine Gruppe mit hohen Ges.Korr.Koeff. gilt : vorausgesetzt, daß diese Korrelationen nicht zufallsbedingt sind, erweisen sich hier die verschiedenartigen Ges.Korr.Koeff. gerade nicht als 'Trennkriterien' in dem Sinne, daß sie Gruppen unterscheiden würden, für welche die WP 'unterschiedlich' funktioniert.

Eine solche Differenzierung z w i s c h e n Gruppen ist - sieht man von der Frage der Zufallsbedingtheit einmal ab - für die +Agg-scores gegeben : für Vpn mit geringen Ges.-Korr.Koeff. zwischen ideal / Wunsch indizieren die

+Agg-scores die ideal-Ebene; für Vpn mit hohen Ges.Korr.-Koeff. zwischen ideal / Wunsch verweisen sie auf die real-Ebene. Wir können diese Werte nicht verallgemeinern : doch legen sie nahe, dieser Frage nach der Rolle der Ges.Korr.-Koeff. als 'Trennkriterien' unter angemesseneren experimentellen Bedingungen gezielter nachzugehen.

Für die Analyse der Bedingungen der differentiellen Validität ist eine - oben bereits erwähnte - Frage besonders wichtig : reagieren alle Vpn, die durch das einheitliche Merkmal einer hohen (geringen) Übereinstimmung zweier Ebenen charakterisiert sind, durchgängig für a l l e Variablen auf e i n e r bestimmten Ebene? Führt etwa eine hohe Übereinstimmung zwischen Wunsch / ideal dazu, daß man die WP durchgehend dem eigenen ideal entsprechend beantwortet : weil man - wie es die Übereinstimmung Wunsch / ideal indizieren könnte - sehr stark Norm-orientiert ist. Führt etwa eine geringe Übereinstimmung zwischen real / Wunsch dazu, daß man seine 'nicht realisierten' Wünsche in die Testreaktionen 'projiziert', daß man diese Wünsche 'substitutiv' befriedigt?

Soweit bei einzelnen Untergruppen für mehrere Variablen eine Korrelation zwischen WP und einer 'Ebene' festgestellt werden konnte, wird diese Annahme einer für alle Variablen einheitlichen Relation nicht bestätigt. Für die Vpn mit geringen Ges.Korr.Koeff. zwischen real / Wunsch z.B. korrelieren die WP-scores - je nach Variable - mit allen 3 Ebenen, - sieht man wieder einmal davon ab, daß diese Korrelationen nicht hinlänglich gesichert sind.

Aus diesem Überblick ist zu entnehmen, daß unser Material für diese Art der Fragestellung unzureichend ist. Nur eine eigene Untersuchung mit gezielten Hypothesen einerseits und mit in irgend einer Form bereits validierten Meßinstrumenten andererseits kann einen Aufschluß darüber geben, ob das Gefüge der 'Ebenen insgesamt gesehen' in einen systematischen Zusammenhang zu bringen ist mit den Testreaktionen. Sollten sich insbesondere Hypothesen darüber ableiten lassen, daß eine hohe allgemeine 'Stabilität' oder 'Instabilität' oder eine hohe bzw. mangelnde 'Angepaßtheit' sich auf bestimmte Weise in der Art der Testreaktion niederschla-

gen, so sind Meßinstrumente nötig, die diesen bestimmten Begriff der 'Stabilität' oder 'Instabilität' bzw. den der 'Angepaßtheit' hinreichend realisieren.

Wenn sich durch eine derartige Untersuchung Trennkriterien nachweisen ließen, welche 'population differences' (Lesser 1959b) für den Test als ganzen beschreiben würden, so wären damit sehr viel günstigere Bedingungen für eine differentielle Validierung gegeben als mit Trennkriterien, die sich nur auf einzelne Variablen bezögen. Die differentielle Validierung eines projektiven Tests verursachte dann den geringsten Aufwand und der Test selbst wäre in der Praxis dann nicht allzu schwierig zu handhaben, wenn er als ganzer für verschiedene Untergruppen validiert werden könnte und wenn nicht für diese Untergruppen ihrerseits die Test-scores für die einzelnen Variablen getrennt zu validieren wären. Das hieße allgemein : durch einen Vortest müsse sich feststellen lassen, nach welchem Auswertungssystem bzw. nach welchen inter-individuellen Normen ein Test 'als ganzer' auszuwerten wäre. Der Aufwand wäre ungleich höher, wenn ein solcher Vortest - im Falle der WP - für jede need-Kategorie getrennt durchzuführen wäre.

Wir können auf Grund unseres Materials nicht entscheiden, ob für die WP ein solches einheitliches Kriterium bestimmt werden kann. Das Material dieser Untersuchung legt indessen nahe, daß die gesuchten Populations-Differenzen nur für jede V a r i a b l e g e t r e n n t festgestellt werden können. So stellten wir etwa fest, daß Vpn mit extremer Selbsteinschätzung auf einer der Ebenen bei e i n e r Variablen hinsichtlich d i e s e r Variablen in einheitlicher Weise auf die WP reagieren. Trennkriterien dieser Art würden z.B. für die Auswertung der Variablen nExh (vgl. Tabelle IX, s.S.244f) zur Folge haben, daß im vorhinein bekannt sein muß, daß die untersuchten Vpn ein besonders starkes Bedürfnis haben 'bewundert zu werden' (hohe Wunsch-Exh-scores) : für diese (und nur für diese) Gruppe wäre dann den unterschiedlichen WP-scores der Hinweis zu entnehmen, daß die Vpn sich in ihren Ideal-Vorstellungen hinsichtlich nExh unterscheiden.

Auf die Probleme, die sich ergeben, wenn derartige Trennkriterien sich nur für jede Variable getrennt bestimmen lassen und auf die Konsequenzen, die in solchem Falle zu ziehen wären, wollen wir erst im Schlußkapitel eingehen. Wir halten noch einmal fest, daß Trennkriterien, die sich auf die Funktionsweise des Tests a l s g a n z e m beziehen, sich nicht in hinreichend gesicherter Weise haben nachweisen lassen.

Wir nehmen daher auch die Werte der Tabelle X nicht in eine zusammenfassende Interpretation der Ergebnisse mit auf. Eine solche Interpretation geben wir im folgenden nur für die Werte der Tabellen IV bis IX.

e) Vergleich der Ergebnisse der verschiedenen Auswertungsschritte

Die Ergebnisse, die sich für einzelne Variablen und einzelne Untergruppen aufgrund der verschiedenen Auswertungsschritte ergaben, beschreiben Relationen zwischen den WP-scores und den Q_s-sort-Daten. Zum einen wurde anhand der Extremgruppenmethode für die WP-scores wie analog für die Q_s-sort-Daten analysiert, ob sie die Vpn i.b. auf ein dichotomisiertes Merkmal differenzieren (Tab. IV - VII, s.S. 227ff); zum anderen wurde für Untergruppen, die nach einem einheitlichen Merkmal zusammengestellt worden waren, untersucht, ob eine korrelative Beziehung zwischen WP-scores und Q_s-sort-Daten nachweisbar ist (Tab. IX, s.S. 242ff).

Die Gegenüberstellung der Ergebnisse nehmen wir im folgenden für jede Variable getrennt vor. Dabei wird sich zeigen lassen, daß die Ergebnisse der Tabelle IX (mit Ausnahme von nAut) jeweils in einen direkten Zusammennang gebracht werden können mit den übrigen Ergebnissen der betreffenden Variablen : so daß es - post festum - gerechtfertigt ist, für diese Werte (der Tab. IX) den weniger strengen Maßstab der einseitigen Signifikanzprüfung gelten zu lassen (vgl. Fußnote 1), S.245). Für die inhaltliche Interpretation der Ergebnisse ist im Auge zu behalten, daß sie

nur für den begrenzten Rahmen der besonderen experimentellen Anordnung vorgenommen und nicht über diese hinaus verallgemeinert werden kann. Erst im Schlußkapitel wird erörtert, welche allgemeineren Schlüsse aus diesen Ergebnissen zu ziehen sind.

n E x h i b i t i o n

Aus Tabelle VII (s.S. 237) geht hervor, daß Vpn mit hohen bzw. geringen WP-Exh-scores sich hinsichtlich ihrer Differenzen zwischen Wunsch-Exh// ideal-Exh sign. voneinander unterscheiden : Vpn mit geringen WP-Exh-scores weisen hohe Differenzen (Wunsch > ideal), Vpn mit hohen WP-scores geringe Differenzen (Wunsch ≈ ideal) auf. Betrachtet man statt der Diff-scores nur die scores e i n e r Ebene, ergibt sich eine vergleichbare, jedoch weniger ausgeprägte (nämlich nur einseitig gesicherte) Relation : Vpn mit geringen (hohen) WP-Exh-scores haben hohe (geringe) Wunsch-scores (vgl. Tab. V, s.S. 231f).

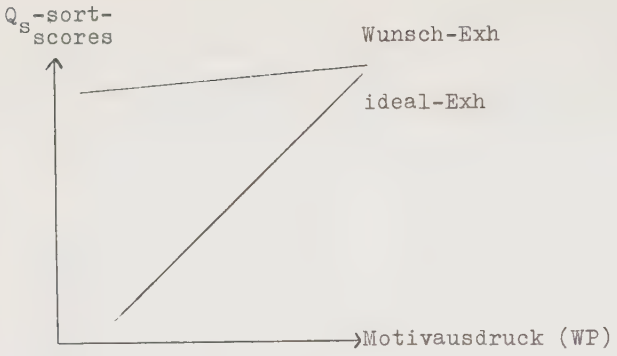
Diesen Ergebnissen zufolge ist eine hohe Einschätzung von nExh auf der Wunsch-Ebene also gepaart mit besonders geringem Motivausdruck in der WP. Diese Relation erscheint zunächst befremdlich, da man umgekehrt nicht einen hohen Motivausdruck bei geringem Wunsch erwartet. Ein weiteres Ergebnis ist hier aufschlußreich : Bei Vpn mit besonders hohem Wunsch-Exh besteht eine K o r r e l a t i o n zwischen Exh WP / ideal-Exh (s.S. 245). Diese Korrelation besagt : je höher - bei hohem Wunsch-Exh - zugleich die Einschätzung derselben Variablen auf der ideal-Ebene ist, desto mehr gelangt das Motiv zum Ausdruck; der starke Wunsch gelangt in der WP nicht zum Ausdruck, wenn nExh im Ich-Ideal negativ bewertet wird. Diesen Zusammenhang kennzeichneten wir dahingehend, daß der Ausdruck des - starken - 'Wunsches' hier durch das 'Ich-Ideal' reguliert würde (s.S. 248).

Dieselbe Relation wird - wenn auch sozusagen in einem gröberen Maßstab - durch den Extremgruppenvergleich beschrieben. Hohe Differenzen Wunsch > ideal und das heißt zugleich ideal< Wunsch sind mit geringem Motivausdruck verknüpft.

Es zeigt sich somit, daß für eine bestimmte Untergruppe jene Relation sehr viel präziser nachweisbar ist, die für die gesamte Versuchsgruppe nur aufgezeigt werden kann, wenn man das Merkmal, welchem die WP-scores zugeordnet werden sollen, dichotomisiert. Diese Präzisierung der Relation zwischen Test-scores und Kriterium wird noch deutlicher, wenn man ein weiteres Ergebnis heranzieht : für die Vpn mit höchsten Wunsch-Exh-scores besteht zugleich eine (negative) Korrelation zwischen Exh WP / Wunsch//ideal : je größer die Differenz, desto geringer die WP-Exh-scores.

Für diese Variable läßt sich somit nachweisen, daß der Motivausdruck in der WP nicht auf eine 'Ebene' allein sondern auf das Verhältnis zweier Ebenen zueinander bezogen werden muß. So ist die Extremgruppen-Differenzierung präziser, wenn man statt der Wunsch-Ebene allein das Verhältnis (die Differenz) zwischen ideal- und Wunsch-Ebene heranzieht. Diese Beziehung zwischen WP-scores und Diff-scores ist wiederum sehr viel enger (nämlich korrelativ), wenn man sie für eine U n t e r gruppe betrachtet.

Durch diese Korrelation Exh WP / Wunsch-Exh//ideal-Exh wird für diese Variable bei einer bestimmten Untergruppe jene Relation beschrieben, die wir bei der Besprechung des approach-avoidance-Konflikt-Modells ausführlich behandelt haben. Dem starken Bedürfnis (approach) wirkt eine hemmende Tendenz entgegen : der Motivausdruck ist um so geringer, je stärker das Motiv im Ich-Ideal abgelehnt wird, bzw. genauer, je stärker diese Ablehnung im Verhältnis zur Bedürfnisstärke ist. Diese Relation läßt sich - analog zum Konflikt-Modell - graphisch darstellen. Die relativ gleichbleibende Motivstärke (Vpn mit höchsten Wunsch-Exh-scores) und die zunehmend größer werdenden Differenzen ergeben hier einen relativ flachen approach- und einen steilen avoidance-Gradienten.



An anderer Stelle wurde bereits dargelegt, daß wir die Terminologie des Konflikt-Modells für dessen Anwendung auf das Testverhalten in Anlehnung an Epstein (1962) präzisieren wollen. Wenn - wie in diesem Falle - nur die Daten der Wunsch- und der ideal-Ebene herangezogen werden, bleibt gerade jene Information unberücksichtigt, die mit den Termini 'approach' und 'avoidance' - wie sie gemeinhin zur Analyse von Testreaktionen verwendet werden - verknüpft wird : Information nämlich über das 'tatsächliche Verhalten'. Die Termini bezeichnen die Alternative, daß ein Motiv (z.B. Aggression) im V e r h a l t e n entweder zum Ausdruck kommt oder nicht. Wir wissen aber aus den obigen Daten nicht, ob die Ablehnung im Ich-Ideal tatsächlich dazu führt, daß das 'Bedürfnis' v e r h a l t e n s m ä ß i g nicht realisiert wird. Die Korrelation beschreibt nur den Sachverhalt, daß das Bedürfnis in der T e s t r e a k t i o n nicht oder verhältnismäßig gering zum Ausdruck gelangt. Wir bezeichnen daher die hemmende Tendenz (in diesem Falle die Ablehnung im Ich-Ideal) als 'inhibition' und die gegenläufige Tendenz als 'expression' (vgl. S.68ff und S.109).

Es sei noch einmal betont, daß diese Beziehung zwischen 'Bedürfnis' und 'Bedürfnisausdruck' nur für eine Untergruppe, nicht für die gesamte Versuchsgruppe nachweisbar ist : nur für die Vpn mit besonders starkem Bedürfnis läßt

sich zeigen, daß der Bedürfnisausdruck danach (inter-individuell gesehen) variiert, wie stark die gleichartige Ablehnung dieses Bedürfnisses ist.

Vpn mit relativ hohen WP-Exh-scores weisen umgekehrt gerade geringe Wunsch-Exh-scores auf (vgl. Tab. V, s.S.231f) : bei einer geringen Einschätzung der Variablen auf der Wunsch-Ebene scheint eine hemmende Tendenz durch das Ich-Ideal nicht wirksam zu sein. Diese Relation wird wiederum deutlicher, wenn man bedenkt, daß für dieselben Vpn mit relativ hohen WP-Exh-scores eine geringe Differenz zwischen Wunsch//ideal gegeben ist (vgl. Tab. VII, s.S.237). Wenn aber sozusagen der Mangel an 'Konflikt' (nämlich geringe Differenzen) zwischen Wunsch und Ich-Ideal mit einem relativ hohen nExh-score in der WP verknüpft ist, ergibt sich eine weitere Problemstellung : es wäre zu prüfen, ob die WP eine relativ hohe 'stimulus-relevance' für nExh besitzt. Diese hohe 'stimulus-relevance' würde dann 'umgangen', wenn die 'Ablehnung' solchen Verhaltens oder solchen Bedürfnisses beesonders hoch ist (also bei Vpn, bei denen hohe Differenzen Wunsch > ideal und geringe WP-scores zusammenfallen, s.o.).

Diesem Problem können wir im Rahmen dieser Arbeit nicht nachgehen; es sei aber auf die ausführliche theoretische Diskussion des Begriffs der 'stimulus-relevance' verwiesen (s.S. 93ff). Diese Diskussion führte zu der Konsequenz, daß die 'stimulus-relevance' definitorisch zu bestimmen und nicht aus den Test-Reaktionen ableitbar sei; daß umgekehrt der Begriff der 'erlebten stimulus-Qualität' nur für das Individuum bestimmt werden könne und so keinen allgemeinen, sondern sozusagen nur einen 'differentiellen' Erklärungswert besitzt. Für die obige Frage nach der Rolle der stimulus-Qualitäten würde das etwa bedeuten, daß Vpn, bei denen weder Wunsch noch Ich-Ideal hinsichtlich nExh besonders ausgeprägt sind, eine Reihe von items in der WP als Hinweisreize für nExh erleben und dementsprechend beantworten; daß aber Vpn mit besonders hohem Wunsch-Exh und besonders starker Ablehnung von nExh im Ich-Ideal diese Hinweisreize für nExh besonders 'sensitiv' wahrnehmen, um sie dann um so stärker zu umgehen (inhibition).

n N u r t u r a n c e

Für diese Variable ließ sich eine k o r r e l a t i v e Beziehung zwischen WP-scores und einem Kriterium nur für jene Gruppen nachweisen, die nach dem Merkmal 'hohe bzw. geringe Ges.Korr.Koeff.' zusammengestellt wurden (vgl. Tab. X, s.S. 262). Die Ergebnisse dieses Auswertungsschrittes jedoch wollten wir von einer zusammenfassenden Interpretation ausschließen, da hier der Einwand einer möglichen Zufallsbedingtheit nicht befriedigend zurückgewiesen werden kann.

Die Extremgruppen-Differenzierung zwischen WP-scores und Kriterium ist für diese Variable besonders gut (nämlich kreuzweise) gesichert. Hohe WP-Nur-scores sind hier mit einer geringen Differenz zwischen real//ideal, geringe WP-scores mit einer hohen Differenz der Richtung ideal> real verknüpft (vgl. Tab. VII, s.S. 237). Dieselbe Relation ergibt sich, wenn man umgekehrt die Vpn mit besonders hohen bzw. geringen Differenzen real//ideal hinsichtlich ihrer WP-scores vergleicht (vgl. Tab. VI, s.S. 235). Betrachtet man beide Ebenen nicht in ihrem Verhältnis zueinander sondern getrennt, so ergibt sich - auch hier wiederum weniger gut gesichert [1] – eine entsprechende Relation für eine der beiden Ebenen : Vpn mit hohen (geringen) WP-Nur-scores weisen zugleich hohe (geringe) real-Exh-scores auf (vgl. Tab. V, s.S. 231f).

Diesen Werten ist zu entnehmen, daß die Testreaktion bei dieser Variablen insbesondere abhängig ist von der Einschätzung des eigenen V e r h a l t e n s : Vpn mit hoher Einschätzung der eigenen 'Hilfsbereitschaft' (real) geben diese 'Hilfsbereitschaft' relativ häufig als Begründung ihrer item-Wahlen in der WP an; diese Begründung aber wird vor allem dann besonders oft genannt, wenn nNur s o w o h l im Verhalten w i e im Ich-Ideal hoch bewertet wird. Bei Vpn dagegen, bei welchen (ihrer Selbstdar-

1)Die Extremgruppen-Differenz ist für die Diff-scores auf dem 2%-, für die real-Nur-scores auf dem 5%-Niveau gesichert (vgl. Tab. V, S.232f und Tab. VII, S.237).

stellung nach) 'Hilfsbereitschaft' im Verhalten besonders gering ausgeprägt ist, wird eine solche 'Gelegenheit zu hilfsbereitem Verhalten' selten als Grund einer positiven item-Wahl in der WP angegeben : und zwar ist hier nNur kaum zu signieren, o b w o h l auch für diese Vpn die Einschätzung von nNur auf der ideal-Ebene hoch ist (hohe Differenz der Richtung ideal> real). Hier läßt sich vermuten, daß diese Vpn bei dieser Variablen das ideal-self zwar der 'a l l g e m e i n e n sozialen Norm' nach beantwortet haben, daß diese Norm jedoch für sie selbst nicht verhaltens-relevant ist. Allein jene Vpn, welche diese Norm zugleich in ihrem Verhalten 'realisieren' (real ≈ ideal), äußern in der WP relativ häufig den Wunsch, 'anderen zu helfen'.

Anders als für die Variable nExh spielt hier also das 'Verhalten' eine Rolle : nur bei Vpn, bei welchen diese 'Eigenschaft' (ihrer Selbstdarstellung nach) in ihrem Verhalten ausgeprägt ist, kommt nNur (vergleichsweise) häufig in der WP zum Ausdruck. Das aber würde bedeuten, daß die WP - anders als Wilde (a.a.O.) annahm - in diesem Falle ein Abbild des 'Seinsbestandes' der Vp geben würde.

n D o m i n a n c e

Diese Variable unterscheidet sich (im Bezugsrahmen unseres Materials) von allen anderen Variablen darin, daß eine k o r r e l a t i v e Beziehung zwischen den WP-scores und Diff-scores für a l l e Vpn nachweisbar ist (vgl. S. 239). Auch hier erweist es sich, daß die Diff-scores in einen engeren Zusammenhang mit den WP-scores gebracht werden können als die scores nur e i n e r Ebene : eine entsprechende Korrelation liegt für die scores einer Ebene nicht vor.

Für diese eine Korrelation zwischen Diff-scores und WP-scores (für alle Vpn) können wir die Frage einer möglichen Zufallsbedingtheit nicht prüfen. Sollte diese Korrelation sich jedoch allein für die Stichprobe der in dieser Untersuchung herangezogenen Vpn aus Zufall ergeben haben, so

ließen sich auch die weiteren Ergebnisse zu dieser Variablen nicht interpretieren : denn sie lassen sich alle aus dieser einen Korrelation erklären und wären somit ebenso zufallsbedingt wie diese selbst.

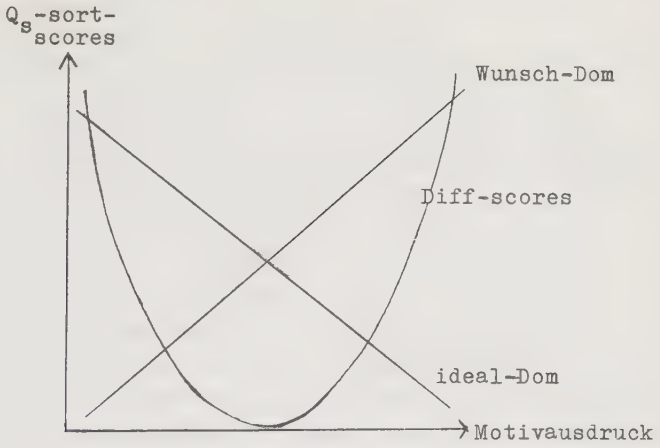
Wir sehen von dieser - nicht geklärten - Frage einmal ab und geben - mit entsprechendem Vorbehalt - auch hier einen Überblick über die Ergebnisse.

Die in dieser Korrelation zwischen WP-scores und Diff-scores (für die Differenz Wunsch//ideal) ausgedrückte Beziehung findet sich - selbstverständlich - wieder in den Ergebnissen des Extremgruppenvergleichs (vgl. Tab. VI, s.S. 235 und Tab. VII, s.S. 237). In weniger (nämlich nicht 'kreuzweise') gesicherter Form ergibt sie sich auch, wenn man für den Extremgruppenvergleich statt der Diff-scores die scores nur der Wunsch-Ebene heranzieht (vgl. Tab. V, s.S. 231f). Die korrelative Beziehung zwischen WP-scores und Diff-scores ergibt sich noch einmal für eine Untergruppe, für die Vpn mit den geringsten ideal-Dom-scores : hier korrelieren die WP-Dom-scores sowohl mit den Wunsch-scores (s.S. 244) wie mit den Diff-scores für Wunsch//ideal (s.S. 250).

Ausschlaggebend für die WP-Reaktion - und zwar gleichartig für die gesamte Versuchsgruppe - ist hier die Einschätzung auf der Wunsch-Ebene : doch ist auch h i e r das 'Bedürfnis' wieder nicht unabhängig vom 'Ich-Ideal' zu sehen. Allerdings hat das Ich-Ideal hier nicht - wie es etwa für die Variable nExh aufgezeigt wurde - die Funktion eines 'Regulativs' : denn der Motivausdruck ist hoch, obwohl die Bewertung dieser 'Eigenschaft' im ideal gering ist; und der Motivausdruck ist gering, obwohl die Einschätzung der Variablen im Ich-Ideal hoch ist.

Die gleichzeitige Rolle der Einschätzungen auf der ideal- und der Wunsch-Ebene für den Motivausdruck läßt sich in Form einer Parabel veranschaulichen, die sich aus der Sub-

straktion von Wunsch- und ideal-scores (Diff-scores) ergibt [1] :



Als allgemeinen Hinweis können wir diesen Ergebnissen zur Variablen nDom - mit dem nötigen Vorbehalt (s.o.) - wieder entnehmen, daß es das 'Gefüge der Ebene' ist, welches den Motivausdruck in der WP bestimmt.

a v o i d a n c e

Bei dieser Variablen ist ein ähnliches Problem gegeben wie für die Variable nDom : auch hier ließ sich eine K o r r e l a t i o n zwischen WP-scores und der Selbsteinschätzung (hier den ideal-avoid-scores) für a l l e Vpn nachweisen, ohne daß geklärt ist, ob es sich um eine zufallsmäßige Signifikanz handelt. Alle übrigen Relationen, die aus dieser einen Korrelation avoid WP / ideal-avoid für alle Vpn (s.S. 225) hervorgehen, sind daher wieder mit Vorbehalt zu interpretieren.

1) Die k o r r e l a t i v e Beziehung zwischen WP-scores und Diff-scores konnte sich ergeben, weil wir von der Richtung der Differenz absahen. Wir gingen aus von der Differenz Wunsch > ideal und werteten die Differenz ideal > Wunsch als negative Differenz der Richtung Wunsch < ideal (vgl. dazu S. 239).

Dieselbe Relation ergibt sich insgesamt für 4 verschiedene Untergruppen (vgl. Tab. IX, s.S. 244f und Tab. X, s.S.262) und - selbstverständlich - für den Extremgruppenvergleich (s.Tab. IV, s.S. 229). Dieser Beziehung zwischen WP-avoid und ideal-avoid wollen wir weiter unten in einem eigenen Abschnitt noch einmal nachgehen (s.S.277ff).

Vermittels Extremgruppenvergleich ließ sich noch eine weitere Relation feststellen : Vpn mit hohem avoid-Ausdruck in der WP weisen hohe Differenzen der Richtung real-avoid Wunsch-avoid auf, während diese Differenz bei Vpn mit geringen WP-avoid-scores gering ist (vgl. Tab. VI, s.S.235 und Tab. VII, s.S. 237). Auf der Wunsch-Ebene wird diese Variable durchgehend negativ eingeschätzt. Demzufolge spielt hier das 'tatsächliche Verhalten' (so wie es erlebt wird) eine Rolle : bei jenen Vpn sind verhältnismäßig viele WP-Antworten mit avoid zu signieren, bei welchen das durch häufige 'Angst' bestimmte Verhalten dem eigenen Wunsch besonders stark entgegengesetzt ist.

n A u t o n o m y

Für das Verhältnis zwischen WP-scores und einem Kriterium liefert das vorliegende Material bei dieser Variablen verhältnismäßig wenig Information.

Einerseits verweist der Extremgruppenvergleich auf eine Beziehung zwischen Wunsch-Ebene und WP-Ausdruck : Vpn mit hohen (geringen) Wunsch-Aut-scores erhalten hohe (geringe) WP-Aut-scores (vgl. Tab. IV, s.S. 229). Diese Relation ist jedoch nicht 'kreuzweise' gesichert.

Andererseits läßt sich für eine Untergruppe (Vpn mit geringsten ideal-Aut-scores) eine K o r r e l a t i o n zwischen WP-scores und real-Aut-scores feststellen (vgl.S.244f und S. 248). Dieser Korrelation nach wäre hier das 'Verhalten' für die Testreaktion ausschlaggebend und zwar um so mehr, je mehr es vom Verhaltensanspruch abweicht. Da jedoch hier keine Zusammenhänge zwischen den verschiedenen Ergeb-

nissen bestehen, muß ihnen ein relativ geringes Gewicht zugesprochen werden. Wir sehen hier von einer weiteren Analyse der Relationen zwischen WP-Aut-scores und den Q_s-sort-Daten ab.

Die Interpretation der Einzel-Ergebnisse, die sich aufgrund verschiedener Auswertungsschritte ergaben, sei hiermit abgeschlossen. Im folgenden wenden wir uns zwei Variablen zu, für deren Interpretation wir zusätzliches Material herangezogen haben.

f) Diskussion der Variablen n Harmavoidance

Für verschiedene Untergruppen wie für die gesamte Versuchsgruppe ergab sich für diese Variable eine Korrelation zwischen den WP-avoid- und den ideal-avoid-scores (vgl. Tab. IX, s.S. 244f und Tab. X, s.S. 262). Wir haben schon darauf hingewiesen, daß diese Korrelation auch für die Untergruppen als zufallsbedingt anzusehen ist, wenn sich nachweisen läßt, daß die Korrelation, die sich für die gesamte Gruppe ergab, a l l e i n für diese besondere Versuchsgruppe und das heißt aus Zufall gegeben ist. Alle Gruppierungen gehörten dann einer Grundgesamtheit an, für welche eine korrelative Beziehung zwischen beiden Meßwerten n i c h t besteht.

Wir können für diese Werte nicht entscheiden, ob sie als nur zufallsbedingt bewertet werden müssen und prüfen daher genauer, ob sich ein Interpretationszusammenhang auf der Grundlage weiterer statistisch gesicherter Relationen ableiten läßt.

An anderer Stelle interpretierten wir die Korrelation zwischen avoid WP / ideal avoid folgendermaßen : die Variable avoid kommt in der WP um so weniger zum Ausdruck, je stärker auf der Ich-Ideal-Ebene ängstliches Verhalten abgelehnt wird; je weniger stark diese Ablehnung ist, um so häufiger sind Test-Antworten mit avoid zu signieren (vgl. S. 248).

Die Einschätzung von avoid auf der ideal-Ebene ist jedoch bei allen Vpn negativ. Warum, so ist zu fragen, kommt die-

se Variable überhaupt so häufig in der WP zum Ausdruck; warum werden bei einigen Vpn ablehnende Reaktionen so oft mit 'Furcht vor Gefahr' etc. begründet? [1] Es ist kaum anzunehmen, daß durch hohe avoid-scores allein angezeigt wird, daß diese Vpn ängstliches Verhalten nicht so extrem negativ im Ich-Ideal bewerten wie andere Vpn.

Eine solche Relation nämlich setzte voraus, daß relativ viele items in der WP von fast allen Vpn als Hinweisreize für avoid erlebt und dann auch entsprechend beantwortet würden, wenn diese Inhalte nicht allzu stark abgelehnt werden. Von Vpn, welche avoid besonders stark ablehnen, würden diese Hinweisreize 'umgangen' : sie reagierten mit weniger avoid als durchschnittlich zu erwarten ist. Demnach wären allein geringe avoid-scores diagnostisch aufschlußreich, indem sie auf 'Vermeidung' verwiesen, während hohe avoid-scores nur die stimulus-Bedingungen des Materials widerspiegelten. [2] Wir wollen im folgenden untersuchen, ob - entgegen dieser Annahme - auch hohen avoid-scores eine diagnostische Relevanz zukommt.

Zur Beantwortung dieser Frage sind die Daten des Erinnerungsexperiments herangezogen worden. In den Untersuchungen zum Wiedererinnern, soweit sie nicht der Analyse des Lernens dienen, sondern vielmehr die 'kognitiven Filter' in Wahrnehmung, Vergessen und Assoziation in ihrer Relation zueinander analysieren, ist folgende Fragestellung von besonderer Relevanz : sind es die positiven oder negativen, die angenehmen oder unangenehmen, die interessierenden oder nicht interessierenden Inhalte, die in der Erinnerung selektiv bewahrt werden (vgl. auch die Arbeiten zum Behalten erledigter oder unerledigter Aufgaben) [3].

Von diesen Arbeiten ausgehend wird hier gefragt, ob der Tatsache, daß bei einem Teil der Vpn gerade die negativen,

---

1) Die Test-scores für avoid variieren zwischen 2% und 30% aller signierten Antworten.
2) Demnach wäre hier der Sonderfall gegeben, daß keine inter-individuellen Unterschiede hinsichtlich der 'erlebten stimulus-Qualität' bestehen (vgl. S. 93ff).
3) für Literaturangaben s.S. 128.

also abgelehnten items der WP vorrangig wiedererinnert werden, eine Bedeutung zukommt. Es wird angenommen, daß durch die bevorzugte Erinnerung der Ablehnungen in der WP ein 'kognitiver Filter' indiziert wird, der vorrangig für negative Aspekte durchlässig ist. Es könnte dies einer besonderen sensitization oder vigilance für Ich-bedrohliche Inhalte entsprechen. Eine solche sensitization aber könnte sich zugleich in der Reaktion auf die W u n s c h - p r o b e dahingehend auswirken, daß besonders Ich-bedrohliche Aspekte an den items wahrgenommen würden, daß von den möglichen 'Seins-Arten', die mit einem item assoziiert werden können, gerade die Existenz-bedrohenden vorrangig selegiert würden, mit der Folge besonders hoher avoid-scores. [1] Ferner ist zu fragen, ob auch der Häufigkeit ablehnender Reaktionen in der WP (also unabhängig von den Variablen betrachtet) eine ähnliche Rolle zukommt. Weiterhin noch kann man annehmen, daß sich eine solche sensitization für Ich-bedrohliche Gegebenheiten auch auf das 'tatsächliche Verhalten' auswirkt; die 'Ängstlichkeit' würde demnach auch im real-self zum Ausdruck gelangen.

Es werden daher alle Vpn in eine Rangreihe danach gebracht, wieviel % der erinnerten Wörter der Kategorie der Ablehnungen in der WP entstammen. Ferner werden Rangreihen der 34 Vpn bzw. 29 Vpn [2] auf der Basis des avoid-score in der WP, des avoid-scores im real-self und der Gesamtzahl negativer Antworten in der WP gebildet. Es seien zunächst alle Korr-Koeff. zusammengestellt : sie müssen dann im einzelnen relativiert werden, da teilweise ein spurious correlation vorliegt.

---

1) Als Beispiel seien hier zum item 'Ozeandampfer' derartige verschiedene 'Seins-Arten' angeführt : "ich wäre Herrscher der Meere", "ich träfe viele Menschen", "ich hätte Angst unterzugehen".
2) Es erhielten nur 29 Vpn eine avoid-score in der WP.

Tab. XI : Relationen zwischen den avoid-scores und selektivem Behalten

Korrelation der Anzahl der aus der neg. Kategorie in der WP erinnerten Wörter (in % der überhaupt erinnerten Wörter) mit :

avoid-scores WP : r_s=0,43 (t=2,48,df=27, $p < 0.01$) [1)]
real-avoid-scores: r_s=0,449 (t=2,84,df=32, $p < 0.05$)

Korrelation der Gesamtzahl negativer Antworten in der WP mit :

real-avoid-scores: r_s=0,451 (t=2,86,df=32, $p < 0.05$)
Anz.d.aus -Ktg. erinnerten Wörter: r_s=0,76 (t=6,76,df=32, $p < 0.001$)

Es zeigt sich also, daß der avoidance-score in der WP um so höher ist, je häufiger jene items erinnert werden, die in der WP negativ beurteilt wurden. Nun korreliert jedoch die Anzahl erinnerter negativer Wörter mit der Anzahl der Ablehnungen in der WP selbst (und zwar hoch sign.), so daß sich die Erinnerung hier nicht allein als Ausdruck eines negativ selegierenden Filters darstellt, sondern auch als Abbild der Verteilung der Wörter, die zu erinnern waren. In dieser Verteilung der zu erinnernden Wörter sind wiederum die mit avoid signierten Ablehnungen enthalten, so daß die Korrelation durch folgende Relation künstlich erhöht wird : je mehr avoid-scores in der WP, desto mehr negative Ablehnungen insgesamt und eine desto stärkere Erinnerung auch eben dieser abgelehnten items. Doch ist die Tatsache eines Zusammenhangs damit noch nicht schon widerlegt : vergleichsweise gibt es ihn nicht für die Agg-Ablehnungen (-Agg), welche einen noch relativ stärkeren Anteil an der Gesamtzahl negativ bewerteter items haben als die avoid-scores.

Daher ist es zusätzlich aufschlußreich, daß die Anzahl der erinnerten negativen items auch mit der von der WP selbst unabhängigen Selbsteinschätzung im Q_s-sort korreliert : je höher das eigene avoid-Verhalten eingeschätzt

1) Die Signifikanz-Prüfungen sind nach Kendall (1948) vorgenommen worden, siehe auch Siegel(1956).

wird, um so größer ist die absolute Zahl der negativen Antworten in der WP und um so höher auch ist der %-Satz an erinnerten abgelehnten items. Wegen des korrelativen Zusammenhangs zwischen Ablehnungen in der WP und Erinnerung dieser Ablehnungen sind auch diese Korrelationen nicht ohne Einschränkung zu interpretieren. Doch ist jedenfalls ein Zusammenhang zwischen der Selbsteinschätzung hinsichtlich avoid und dem Ausmaß ablehnender Reaktionen in der WP nachgewiesen. [1)]

Es kann diesen Werten entnommen werden, daß zwischen dem Grad der Selbsteinschätzung hinsichtlich angst-bestimmten Verhaltens (real-avoid-scores) und der bevorzugten Erinnerung gerade der abgelehnten items, wie auch dem Ausmaß ablehnender Reaktionen in der WP selbst ein sign. Zusammenhang besteht. Ebenso zeigt sich eine derartige Relation zwischen dem Ausdruck von avoid in der WP und der Anzahl der aus der Erinnerung reproduzierten negativ bewerteten items der WP.

Für jene Vpn, bei welchen sowohl der avoid-score in der WP, wie der real-avoid-score relativ hoch sind und welche zugleich vorrangig aus der WP die abgelehnten items wiedererinnern, wie auch besonders häufig ablehnend auf die WP-items reagieren, liegt der Schluß nahe, daß hier durchgängig eine sensitization für 'negative' Inhalte zum Ausdruck kommt. Diese zeigt sich darin, daß diese Vpn - ihrer Selbsteinschätzung nach - die in den avoid-items des Q-sort's dargestellten Situationen häufiger als andere Vpn als bedrohlich erleben; daß die items der WP bei ihnen besonders häufig bedrohliche Assoziationen auslösen ('zertreten', 'zerstört' werden), daß die WP-items vorwiegend ablehnende Reaktionen hervorrufen und daß vorwiegend abgelehnte WP-items im Gedächtnis haften bleiben. [2)]

---

1) Hier liegt keine spurious correlation vor, weil zwischen den real-avoid- und den avoid-scores in der WP, welche ja die Gesamtzahl negativer Antworten mitbestimmten, keine Korrelation besteht.

2) Zwischen der Anzahl ablehnender Reaktionen auf die WP-items und der Anzahl der wiedererinnerten abgelehnten WP-items besteht selbstverständlich eine 'unechte' Korrelation; es werden diese Reaktionstendenzen jedoch beide angeführt, da sie mit dem unabhängigen Kriterium der Selbsteinschätzung im Q-sort (real-avoid) korrelieren.

Diesem Material kann also entnommen werden, daß nicht allein das Ich-Ideal den Ausdruck von avoid 'reguliert', daß nicht allein mit zunehmender negativer Bewertung dieser Variable im Ich-Ideal um so weniger avoid-Ausdruck in die WP gelangt. Vielmehr ist ein hoher avoid-score nicht nur dahingehend zu verstehen, daß hier etwa die aufgrund der 'erlebten stimulus-Qualität' nahegelegten avoid-Reaktionen nicht wirksam umgangen worden wären; sondern ein hoher avoid-score in der WP steht positiv in Zusammenhang mit einer besonderen Reaktionsbereitschaft : diese drückt sich in einer allgemeinen 'sensitization' für ich-bedrohliche Inhalte (WP) und ich-bedrohliche Situationen (real-avoid), wie im vorrangigen Behalten abgelehnter items aus.

Unserem Material nach scheint es, daß es sich hier um eine relativ allgemeine Relation handelt. Trennkriterien - soweit sie jedenfalls anhand des Ebenen-Konzeptes gebildet werden können - mußten bei dieser Variablen nicht eingeführt werden. Für die Korrelation der verschiedenen avoid-scores (aus WP und Q_s-sort) mit den Daten des Erinnerungsexperiments wurden alle Vpn herangezogen. Auch aus diesen Korrelationen können wir nicht allgemeine Schlußfolgerungen ziehen, die hinreichend gesichert wären. Doch läßt sich ihnen die Hypothese entnehmen, daß den avoid-Reaktionen in der WP eine hohe diagnostische Valenz zukommt. Hier hätte eine genauere Untersuchung einzusetzen.

g) Diskussion der Variablen n Aggression

Für diese Variable ergibt sich, ebenso wie für die unter Abschnitt e) diskutierten Variablen, ein Leitfaden für die Interpretation, wenn man die verschiedenen Ergebnisse aufeinander bezieht.

Nach Tabelle IV (s.S. 229) beantworten Vpn, die ihr Verhalten als nicht-aggressiv einschätzen, die items der WP besonders häufig mit der Ablehnung von Aggressivität (-Agg); während bei Vpn, welche ihr Verhalten - im Vergleich zur Versuchsgruppe - als aggressiv beschreiben, die Ablehnung von Aggression relativ selten zu signieren ist. Dieser Extremgruppen-Differenz nach sind es in-

aggressive Vpn, die eine Ablehnung von Aggressivität besonders häufig als Grund einer negativen item-Wahl angeben.

Einem weiteren Ergebnis zufolge ist dieser Schluß jedoch in entscheidender Weise zu modifizieren : es sind Vpn mit hohen Differenzen der Richtung Wunsch-Agg > real-Agg, welche einen hohen -Agg-scores erhalten; eine relativ geringe Ablehnung von nAgg findet sich dagegen bei den Vpn mit einer hohen Differenz der Richtung real > Wunsch. Dieselbe Relation ergibt sich, allerdings einseitig gesichert, auch für den Quotienten $^{-Agg}/_{+Agg}$ (vgl. Tab. VI, s.S. 235).

Demnach ist die häufige Ablehnung von nAgg in der WP nicht allein mit in-aggressivem Verhalten in Zusammenhang zu bringen, sondern zugleich mit einem starken und insbesondere vom Verhalten stark abweichenden Aggressions-Bedürfnis. Umgekehrt findet sich die Ablehnung von nAgg verhältnismäßig selten bei Vpn, die ihr Aggressions-Bedürfnis als gering, ihr Verhalten jedoch als stark aggressiv angeben.

Bezieht man gleichzeitig die positiven Aggressions-Antworten mit ein, indem man den Quotienten $^{-Agg}/_{+Agg}$ betrachtet (s.o.), so läßt sich der Zusammenhang noch weiter präzisieren : Vpn mit hohen Wunsch- aber geringen real-Agg-scores geben sehr viel mehr -Agg-Antworten als +Agg-Antworten; Vpn, die hohe real-, aber geringe Wunsch-Agg-scores aufweisen, antworten vergleichsweise häufiger mit +Agg (denn bei diesen ist der Quotient sehr viel kleiner als bei obigen Vpn). Warum aber reagieren Vpn mit der Differenz real Wunsch häufiger mit +Agg als Vpn mit der Differenz Wunsch real? Wie läßt sich erklären, daß Vpn mit hohem Agg-Bedürfnis (self-report) in der WP sehr viel häufiger Aggression ablehnen, denn als Bedingung für eine positive Wahl angeben?

Diese Relation wird noch durch ein weiteres Ergebnis gestützt. Für die Vpn mit besonders geringen real-Agg-scores (s.o.) ergab sich die Korrelation $^{-Agg}/_{+Agg}$ WP / Wunsch-Agg (vgl. S. 245). Diese Korrelation besagt : Je größer das Aggressions-Bedürfnis bei geringer verhaltensmäßiger Aggressivität ist, desto stärker ist die Ablehnung

von Aggression in der WP im Verhältnis zu den positiven Aggressions-Antworten; je geringer die Wunsch-Agg-scores sind, desto häufiger sind die WP-Antworten mit +Agg zu signieren.

Diese mehrfach bestätigte Relation, daß hohe Wunsch-Agg-scores zusammenfallen mit hohen -Agg-scores, während bei Vpn mit geringen Wunsch-Agg-scores relativ hohe +Agg-scores in der WP festzustellen sind [1], wollen wir im folgenden durch verschiedene Hypothesen auf ihren Bedingungszusammenhang hin untersuchen. Es handelt sich dabei um Interpretations-Hypothesen, anhand welcher aus dem vorliegenden Material zusätzliche Information zur Klärung der obigen Relation zu gewinnen ist.

Das Zusammenfallen von hohen Agg-Wünschen (self-report) und häufiger Ablehnung von nAgg in der WP läßt auf einen Zusammenhang schließen, den wir an früherer Stelle schon einmal diskutiert haben (vgl. S. 79) : ein starkes Bedürfnis führt zu einer 'sensitization' gegenüber aggressiven cues, so daß die WP-items besonders häufig unter dieser 'Wahrnehmungshypothese' aufgefaßt werden. Die auf diese Weise mit den WP-items assoziierten aggressiven Inhalte aber werden (weil sie sozial oder subjektiv nicht akzeptabel sind) im nachhinein a b g e w e h r t : Aggression kommt zwar in eben dem Maße zum Ausdruck, wie sie mit den WP-items assoziiert wurde, doch erscheint sie in 'negativer' Form, nämlich in Form einer 'thematischen Vermeidung'.[2]

Diese Hypothese ist stichhaltig nur für den Fall, daß es sich um 'nicht zugelassene' bzw. um 'konfliktbesetzte' Agg-Wünsche handelt : nur für diesen Fall ist zu erwarten, daß es sich bei den hohen -Agg-scores um d e f e n s i v e Reaktionen handelt.

---

1) Es ist zu beachten, daß diese Relation nicht für die Wunsch-scores selbst aufgewiesen wurde, sondern feststellbar ist, wenn man gleichzeitig Wunsch- und real-Agg-scores betrachtet.

2) Dieser Begriff wurde bei der Diskussion der Lesser'schen Arbeiten zum TAT eingeführt (vgl.S. 68). Während der übliche Begriff von 'avoidance' die vollständige Vermeidung von Motivausdruck bezeichnet, soll der Begriff 'thematische Vermeidung' den Sachverhalt bezeichnen, daß die 'Vermeidung' (Ablehnung) in Form einer negativen Bewertung s p r a c h l i c h ausgedrückt wird.

Es sind daher Anhaltspunkte zu gewinnen, welche die Annahme einer Verteidigungsreaktion bestätigen könnten. Hier ist auf die obige Diskussion der Variablen avoid zu verweisen : Verschiedene Korrelationen zwischen avoid-scores und den Daten hinsichtlich selektiven Vergessens werteten wir als Hinweis darauf, daß mit hohen real-avoid-scores eine hohe sensitization gegenüber ich-bedrohlichen cues verbunden sei, eine allgemeine Bereitschaft, auf 'negative Inhalte' anzusprechen (s.S. 281). Es wäre denkbar, daß hohe -Agg-scores (bzw. hohe $^{-Agg}/_{+Agg}$-scores) dadurch zustandekommen, daß diese Vpn besonders sensitiv auf 'nicht zugelassene' Inhalte reagieren, - um sie dann allerdings 'thematisch' abzuwehren.

Wir prüfen daher für die Vpn mit den höchsten -Agg- bzw. $^{-Agg}/_{+Agg}$-scores, ob eine Relation zwischen diesen WP-Agg-scores und den real-avoid-scores nachweisbar ist.

Es ergeben sich folgende Korrelationen (bei N = 10) :

$^{-Agg}/_{+Agg}$ WP / real-avoid :	r_s=+0,586	$p < 0.05$
-Agg WP / real-avoid :	r_s=+0,553	sign. bei r_s=0,564

Diese Korrelationen lassen sich dahingehend interpretieren, daß zunehmend hohe -Agg-scores verknüpft sind mit einer zunehmenden Sensitivierung gegenüber Ich-bedrohlichen cues. Gleichwohl können wir diese Korrelation nicht schon als Beweis für die obige Hypothese betrachten. Das Kriterium (real-avoid) selbst zeichnet sich zwar durch eine gewisse construct-validity aus, sofern man die Dimensionen 'Sensitivität in der Wahrnehmung' und 'Selektivität des Behaltens' definitorisch miteinander verknüpft (vgl. S.279). Doch wären eine Reihe von Gegenannahmen zu prüfen, bevor wir aus der obigen Relation Verallgemeinerungen herleiten können.

So müßte sich - in Umkehrung der obigen Relation - nachweisen lassen, daß hohe +Agg-scores mit einer geringen 'defensiveness' bzw. 'sensitization' verknüpft sind und daß hohe +Agg-scores gerade dann auftreten, wenn diese Variable nicht 'konfliktbesetzt' ist, d.h. wenn den hohen

Agg-Wünschen ein ebenso aggressives Verhalten (real-Agg) entspricht [1]. Beide Annahmen lassen sich für unser Material nicht bestätigen.

Die Vermutung, daß besonders hohe -Agg-scores (bzw. $-Agg/+Agg$-scores) auf sensitization und defense zurückzuführen sind, ist demnach nicht hinreichend geklärt. Da mit dieser Vermutung aber eine entscheidende Frage in bezug auf die Funktionsweise der WP angeschnitten ist, sollte sie in Form einer Hypothese in speziell auf dieses Problem ausgerichteten Experimenten analysiert werden.

Noch eine weitere Frage kann an die korrelative Beziehung zwischen -Agg-scores und real-avoid-scores geknüpft werden, jedoch gleichfalls auf der Grundlage unseres Materials nicht entschieden werden. Sie sei kurz beschrieben, weil sie auf Probleme verweist, die mit der S i g n i e r u n g der Test-Antworten verbunden sind, Probleme, von denen wir bei der Auswertung des Materials bisher abgesehen haben, welche jedoch bei gezielteren Validitätsuntersuchungen nicht unberücksichtigt bleiben dürfen.

Bei der Besprechung unseres Auswertungsschemas verwiesen wir darauf, daß eine strikt formale Signierung des rein sprachlichen Gehalts der Test-Antworten häufig kaum zu leisten ist, da die oft nur stichwortartigen Begründungen in den Testprotokollen durch den Auswerter in irgendeiner Form 'sinngemäß' erweitert werden müssen, sollen sie sich überhaupt kategorisieren lassen (vgl.S. 194f). Es könnte auch - wie gleich zu zeigen sein wird - diese mangelnde Genauigkeit in der Signierung sein, auf welche die obige Korrelation verweist.

Die häufigen ablehnenden Reaktionen (-Agg) könnten darauf zurückzuführen sein, daß die Vpn die items weniger als cues erleben, welche Aggression 'gegenüber anderen' beinhalten, sondern als Hinweisreize für eine eigene Ich-Bedrohung. Das würde bedeuten, daß in den -Agg-Antworten

---

1) Statt der Differenz Wunsch-Agg > real-Agg müßte also gegeben sein : Wunsch-Agg ≈ real-Agg bei hohen Wunsch- und real-Agg-scores.

nicht zum Ausdruck kommt, daß man Aggressivität anderen gegenüber ablehnt, sondern daß diese ablehnenden Reaktionen darauf verweisen, daß die Vpn s e l b s t nicht bedroht werden möchten. Die Vp brächte also nicht zum Ausdruck, daß sie für andere keine Gefahr bedeuten möchte, sondern ihre Antwort indizierte, daß sie selbst keiner Gefahr ausgesetzt sein möchte. In der Tat sind die Test-Antworten in dieser Hinsicht häufig mehrdeutig : es läßt sich nicht immer entscheiden, ob die Begründung einer negativen Wahl als 'Projektion' der eigenen 'Angst vor etwas' zu verstehen ist oder als Ausdruck dessen, daß man anderen keinen Schaden zufügen möchte.

Als Beispiel sei eine Antwort zum item 'Eisberg' angeführt : "der ist zu gefährlich". Diese Kurzantwort läßt sich inhaltlich in zweifacher Weise ergänzen :

a) "Ich möchte für andere nicht so gefährlich sein wie ein Eisberg" : -Agg
b) "Das Leben als Eisberg ist mir zu gefährlich" : avoid

Im ersteren Falle kommt zum Ausdruck, daß man für andere nicht gefährlich sein möchte; im Falle b) wird indiziert, daß man selbst es ist, der sich bedroht fühlt.

Eine derartige qualitative Analyse der Testreaktionen haben wir uns nicht in dieser Arbeit zur Aufgabe gemacht. Im Gegenteil war es unser Ziel, eine qualitativ-diagnostische Interpretation der Testprotokolle zu vermeiden und allein eine sprachlich-formale Kategorisierung vorzunehmen : allein der sprachlich manifeste (nicht der erschlossene 'latente') Gehalt der Test-Antworten war darauf zu untersuchen, ob er sich einem Kriterium zuordnen läßt. Diese probeweise vorgenommene Auswertung der WP wäre ihrerseits einer eingehenden Analyse zu unterziehen : die Validierung eines Tests ist zugleich die Validierung des verwendeten Auswertungssystems. Alle Ergebnisse dieser Untersuchung sind an die Eigenart des benutzten Auswertungsschemas gebunden : es ist nicht ausgeschlossen, daß eine andere Form der Signierung zu einer engeren Beziehung zwischen Index und Indiziertem führen könnte.

Wir kehren zu unserer Ausgangsfrage zurück. Ungeklärt blieb bisher, warum hohe +Agg-scores (bzw. geringe $-Agg_{/+Agg}$-scores) zusammenfallen mit geringen Wunsch-Agg-scores bzw. mit hohen Differenzen der Richtung real > Wunsch.

Diese Relation könnte dahingehend verstanden werden, daß in den hohen +Agg-scores das relativ stark aggressive Verhalten (real) zum Ausdruck kommt, - ungeachtet dessen, daß dieses Verhalten wunschmäßig abgelehnt wird. Warum aber reagieren Vpn auf die Instruktion, sich anhand der items etwas zu 'wünschen', mit positiven Agg-Antworten, wenn sie doch eben diese Aggression 'wunschmäßig' (dem self-report nach) ablehnen?

Hier läßt sich die Hypothese formulieren, daß die WP-Reaktion auf die 'wahren' Aggressions-Wünsche verweisen, während den Q_s-sort-Daten der Hinweis auf eine mangelnde 'Einsicht' in diese Wünsche zu entnehmen sei. Es läßt sich jedoch auch annehmen, daß im self-report zum Ausdruck gebracht wird, daß man sich 'wünschte', n i c h t aggressiv zu sein, während in der Testreaktion die tatsächlich vorhandene Aggressions-Tendenz 'ungesteuert' in Erscheinung tritt, da sie hier - anders als im Verhalten - nicht zu sozialen Komplikationen führt. [1] Demnach würde in der Testreaktion ein Bedürfnis 'befriedigt', das zwar auch im overt behavior (real) einen 'Ausdruckskanal' findet, bei seiner 'verhaltensmäßigen Realisation' jedoch zu Konflikten führt (nämlich real-Agg> Wunsch-Agg) [2].

---

1) Hier ist auf das Problem zu verweisen, daß die Wunsch-Ebene einerseits den 'Antrieb' mißt, andererseits aber auch den 'Wunsch', daß die eigene Antriebsseite anders geartet sein möge, - etwa der Wunsch (nicht das Ideal), dominant zu sein, obwohl der Antrieb, sich unterzuordnen, nachzugeben etc. stärker ist. Zwar wird in der Instruktion gefordert, man solle angeben, wie man sich verhalten würde, wenn die ä u ß e r e n Bedingungen eine freie Wahl zuließen; diese Aufforderung impliziert nicht, daß man sich sozusagen auch die 'i n n e r e n Bedingungen' wählen könne, also etwa den Wunsch äußern könne, wie man sich gerne verhalten möchte, hätte man eine andere 'Antriebsstruktur'. Hier werden die Schwierigkeiten einer experimentellen Realisation der Wunsch-Ebene besonders deutlich.

2) vgl. dazu die Beschreibung des 'alternative channel model', S. 55 ff.

Warum aber wird nAgg bei diesen Vpn auf der Wunsch-Ebene besonders negativ eingeschätzt? Warum wird der Wunsch, Aggressionen 'auszuleben', so ausdrücklich abgelehnt? Wir prüfen auch hier - mit dem nötigen Vorbehalt gegenüber der Annahme, daß das verwendete Kriterium (real-avoid) 'gültig' sei - , ob sich die besonders negativen Einschätzungen von nAgg auf der Wunsch-Ebene in Zusammenhang bringen lassen mit einer besonders sensitiven Reaktion auf Ich-bedrohliche Inhalte. Handelte es sich bei dieser negativen Einschätzung von nAgg um defensives Verhalten, so wäre bestätigt, daß bei diesen Vpn Aggressivität konfliktbesetzt ist.

Wir prüfen daher für die Vpn mit den geringsten Wunsch-Agg-scores, ob sich zwischen diesen Wunsch-Agg-scores und den real-avoid-scores eine Beziehung nachweisen läßt. Es ergibt sich die Korrelation (bei N = 13) :

$$\text{Wunsch Agg / real-avoid} \quad : \quad r_s = -0{,}53 \quad p < 0.05$$

Folgende Relation wird angezeigt : je stärker die Vpn sensitiv auf Ich-bedrohliche cues reagieren (real-avoid), desto negativer wird nAgg auf der Wunsch-Ebene eingeschätzt. Demnach erscheint die obige Annahme bestätigt. Trotzdem wollen wir auch hier keine a l l g e m e i n e Schlußfolgerung ziehen : das verwendete Kriterium (real-avoid) erscheint uns nicht hinlänglich gesichert. [1] Wir folgern nur, daß durch unser Material nahegelegt wird, die obige Hypothese eingehender, d.h. durch angemessenere Mittel, zu prüfen.

Es seien noch einmal die obigen Ausgangswerte wiederholt und die Argumentation zusammengefaßt :

Vpn mit hoher Differenz Wunsch-Agg > real-Agg weisen hohe $^{-Agg}/_{+Agg}$-scores auf,

Vpn mit hoher Differenz real-Agg > Wunsch-Agg weisen geringe $^{-Agg}/_{+Agg}$-scores auf.

---

1) Hier ist auf das Problem zu verweisen, daß die Inter-Relationen zwischen den verschiedenen Variablen des Q_s-sort-Verfahrens bekannt sein müssen, sollen sich die Meßwerte eindeutig interpretieren lassen (vgl.S.171f).

Sollten die obigen Annahmen zutreffen, so handelte es sich bei den Vpn mit Wunsch-Agg > real-Agg um besonders 'stark gesteuerte' Vpn : sie sind nicht aggressiv, obwohl sie 'es gerne wären'; sie assoziieren mit den WP-items besonders häufig aggressive Inhalte, wehren diese jedoch ab, indem sie die betreffenden items ablehnen ('thematische Vermeidung'). Die Vpn mit real-Agg > Wunsch-Agg sind vergleichsweise 'wenig gesteuert' : sie s i n d aggressiv, obwohl sie 'es lieber nicht wären'; und sie bringen ihre aggressiven Tendenzen in der WP - vergleichsweise häufig - ungehindert zum Ausdruck.

Es sei noch einmal betont, daß diese Schlußfolgerungen aus den Korrelationen zwischen nAgg- und real-avoid-scores nur einen hypothetischen Charakter haben : es handelt sich nicht um Erklärungen, die sich auf einen 'Beweis' stützen würden. Durch die Tatsache jedoch, daß sich ein Interpretations-Zusammenhang herstellen läßt, wenn man die Daten in den Begriffen des Ebenen-Konzeptes analysiert, sehen wir uns in unserer Grundannahme bestärkt :

Die 3 Ebenen, Antrieb, seine Realisation und seine Kontrolle können nicht getrennt betrachtet werden; keine Ebene ist für sich genommen aufschlußreich. Die T e s t - reaktion ist als Resultante dieser 3 Ebenen zu verstehen; das bedeutet : das G e f ü g e der Ebenen ist in Betracht zu ziehen, sollen Testreaktionen, soweit sie sich als 'needs' beschreiben lassen, in ihren Bedingungen analysiert werden.

Wir schließen unsere Interpretation der einzelnen Ergebnisse mit dem Hinweis auf jenes Beispiel, das den Anstoß zu dieser Untersuchung gab. Einleitend zitierten wir einen Katalog verschiedener Interpretationsmöglichkeiten, den Hörmann (1964a) für die WP-Antwort 'weil der so schön beißen kann' (positive Wahl des items 'Tiger') zusammenstellte (vgl. S. 26) :

> "daß die Vp aggressiv ist und sein will, daß sie nicht aggressiv ist, aber aggressiv sein möchte, oder daß sie nicht aggressiv ist und auch bewußt nicht aggressiv sein möchte, ihren unbewußten Wunsch nach Aggressiv-Sein aber hier projiziert, oder daß sie aggressiv ist, aber bewußt nicht aggressiv sein möchte und deshalb hier ihr Aggressiv-Sein-Wollen projiziert?"

Anhand unseres empirischen Materials ließ sich bestätigen, daß in der Tat mit derart vielfältigen Relationen zwischen Test-Index und Indiziertem zu rechnen ist, daß eine Testreaktion sehr verschiedenartigen psychischen Bedingungen zugeordnet werden kann. Es wurde deutlich, daß sich eben diese psychischen Bedingungen systematisieren lassen, wenn man sie in den Begriffen eines Ebenen-Konzeptes, wie es hier vorgeschlagen wurde, analysiert.

Im folgenden bleibt noch einmal genauer zu untersuchen, welche a l l g e m e i n e n Schlußfolgerungen aus den Ergebnissen dieses Erkundungsversuchs zu ziehen sind, welcherart Konsequenzen sich für das Problem der Funktionsweise eines projektiven Tests ergeben.

S c h l u ß b e t r a c h t u n g

1. Zusammenfassender Überblick über den Gang der Argumentation

Wenn es unsere Aufgabe ist, noch einmal zusammenfassend zu untersuchen, welche a l l g e m e i n e n Schlußfolgerungen sich den Ergebnissen dieses Erkundungsversuchs entnehmen lassen, so sollen zunächst noch einmal die Fragestellung wie die verschiedenen Schritte, welche zu dem vorliegenden Experiment führten, kurz rekapituliert werden. [1)]

Ausgangspunkt der Arbeit war, daß die Projektionshypothese nur einen geringen Erklärungswert besitzt für die Frage, wie eine Testreaktion zustandekommt und für welcherart Phänomen sie als Indiz gewertet werden kann. In der Literatur zu projektiven Tests ist seit langem unumstritten, daß es nicht allein die 'verborgenen' oder 'latenten', die vor dem Einblick der Umwelt 'gehüteten' oder auch die 'unbewußten' Inhalte sind, welche in die Testantworten 'hinausverlegt' werden. Es ist unbestritten, daß ein Testprotokoll nicht - trotz seiner scheinbar unmittelbaren 'Aussagekraft' - als ein direkter 'Schlüssel zur Seele' verstanden werden kann.

Anhand eines Ausschnitts aus der vorliegenden Literatur ließ sich zeigen, daß eine Fülle von verschiedenartigen Beziehungen zwischen Index und Indiziertem eine einfache Punkt-zu-Punkt-Zuordnung von Test-score und Kriterium unmöglich machen. Dieser Tatbestand führte uns zu folgender Überlegung : die Mehrdeutigkeit, die einem Test-score anhaftet, kann nur dann durch einzelne exakt bestimmbare Relationen zwischen Testreaktion und einem durch diese Reak-

1) Dieses Kapitel, als zusammenfassendes Abschlußkapitel, kann nicht die Aufgabe haben, durch Querverweise etc. auf den gesamten übrigen Text hinzuweisen. Es geht hier vielmehr darum, noch einmal in geraffter und thetischer Form das Grundproblem der Arbeit kurz und zusammenhängend, unter Verzicht auf Begriffserklärungen und theoretische Ableitungen, zu skizzieren.

tion indizierten psychologischen Sachverhalt ersetzt werden, wenn man den Test-score, welcher nur sozusagen das Endprodukt eines Reaktionsablaufs darstellt, von eben dieser Reaktion nicht abstrahiert betrachtet. Ein Test-scores ist nicht nur Hinweis a u f Verhalten, sondern Ausdruck v o n Verhalten (vgl. Loevinger 1965).

Diese Vielfalt aufweisbarer Relationen zwischen Index und Indiziertem legte eine weitere Konsequenz nahe : die Gültigkeit eines projektiven Verfahrens läßt sich nur d i f f e r e n t i e l l , d.h. getrennt für verschiedene Gruppen, bestimmen. Somit war das Problem gestellt, T r e n n k r i t e r i e n zu gewinnen, anhand welcher diese 'population differences' erfaßt werden könnten.

Ein weiterer Ansatzpunkt ergab sich aus der Überlegung, daß dem Mißerfolg aller Versuche einer empirischen Validierung projektiver Tests nur durch eine Orientierung an jenem Validierungs-Modell zu begegnen ist, nach welchem die Gültigkeitsbestimmung eines Verfahrens die Gültigkeit des verwendeten Kriteriums voraussetzt. Nach dem Modell der construct-Validität ist zunächst auf der Ebene der Theorie (nicht der der einfachen empirischen Zuordnung) zu analysieren, welcher theoretische Begriff (construct) durch das verwendete Verfahren 'realisiert' wird. Ein Netzwerk von Relationen zwischen Test-scores und Meßwerten verschiedener, dem Test theoretisch zugeordneter Dimensionen liefert das empirische Abbild der theoretischen Definition jener Realität, für welche der Test valide ist.

Es war nicht die Aufgabe der vorliegenden Untersuchung, eine derartige Gültigkeitsbestimmung vorzunehmen, ein solches Netzwerk in allen seinen Bestandteilen experimentell herzustellen. Die Analyse der Funktionsweise eines projektiven Tests war jedoch in einem entscheidenden Punkt am Modell der construct-validity zu orientieren : an der Forderung, sowohl die T e s t reaktion wie jenes Verhalten, von welchem man annimmt, daß es durch den Test gemessen wird, auf ihre Bedingungen hin und insbesondere auf ihre gemeinsamen Bedingungen hin zu analysieren.

Die Ergebnisse vorliegender Experimente zu projektiven Verfahren einerseits und andererseits die Forderung, Test u n d Kriterium auf einer gemeinsamen Basis zu analysieren, führte uns zu der Annahme, daß ein p e r s ö n l i c h k e i t s -theoretisches Begriffssystem für die Aufgabe einer Test-Analyse heuristisch am fruchtbarsten sei. Es war ein Begriffsrahmen zu finden, innerhalb dessen sich Index und Indiziertes in ihren Beziehungen beschreiben und welchem sich zugleich die verschiedenen allgemein-psychologischen Modelle subsumieren ließen, in denen bisher die Kenntnisse über die Funktionsweise projektiver Verfahren beschrieben worden sind.

Das Ebenen-Konzept erfüllt diese Aufgabe nicht ganz : die allgemein-psychologisch gefaßten Mechanismen, wie sensitization und defense, coping - avoiding, revealing - conseiling, werden nicht in das Gebäude einer geschlossenen Theorie integriert, sondern eklektisch als Erklärungsprinzipien hinzugezogen. Es ließ sich aber zeigen, daß mit Hilfe des Ebenen-Konzeptes Bedingungen solcher Mechanismen wie sensitization oder defense beschrieben werden können : So muß nicht ausgeschlossen werden, daß eine solche Integration eines allgemein-psychologischen und eines persönlichkeitstheoretischen Begriffssystems prinzipiell möglich ist.

In dieser Untersuchung blieben - zugunsten persönlichkeitsspezifischer Merkmale - allgemein-psychologische Fragestellungen völlig ausgeklammert. So wurde nicht untersucht, ob die Testreaktion auch mit allgemeinen Reaktionstendenzen, etwa mit verschiedenen 'cognitive response styles' in Zusammenhang zu bringen ist. Ebenso blieben alle situativen, d.h. durch die Test- S i t u a t i o n bedingten Determinanten des Testverhaltens unberücksichtigt. Schließlich sehen wir in dieser Untersuchung von jenen Bedingungen ab, die durch den Test selbst gegeben sind : eine Analyse der stimulus-Bedingungen wurde nicht vorgenommen. Stattdessen versuchten wir aufzuzeigen, daß der Begriff der 'stimulus-relevance' der 'experimentellen Ökologie' zuzurechnen ist und daher nur d e f i n i t o r i s c h bestimmbar ist. Der entsprechende 'Lebensraumbegriff', die 'erlebte stimulus-Qualität', kann bei einem vorgegebenen Verfahren wie der Wunschprobe nur im Vollzuge

der (differentiellen) Validierung selbst, d.h. implizit bzw. im nachhinein bestimmt werden.

Das Ebenen-Konzept, das wir für den Zweck dieser Untersuchung einführten, wurde als eine Art Kompromiß aus dem Ansatz von Rosenzweig und jenem von Leary hergeleitet. Ein methodisches Problem bestand in der Aufgabe, dieses theoretische Konzept experimentell zu 'r e a l i s i e r e n' .

Die Q_s-sort-Methode erwies sich aus technischen Gründen als besonders geeignet : sie ist indessen unzureichend, weil eine Gültigkeitsbestimmung fehlt. Wir stellten daher ersatzweise eine Reihe von Hypothesen auf, deren Bestätigung als Hinweis für eine gewisse construct-validity des Verfahrens bewertet werden kann.

Diese Hypothesen bezogen sich jedoch allein darauf, daß mit dem Q_s-sort-Verfahren verschiedene 'E b e n e n' gemessen würden : sie bestätigen nicht, daß wir tatsächlich ein Meßinstrument etwa für nDominance oder nAggression besitzen.

Für diese Frage, wieweit mit unserem Instrumentarium einzelne V a r i a b l e n getroffen werden, haben wir einen einzigen Anhaltspunkt : wir wollten das Q_s-sort-Verfahren wie das Auswertungsschema für die Wunschprobe dann - mit Einschränkung - als brauchbare Instrumente ansehen, wenn sich zwischen den Daten beider Verfahren - den Erwartungen entsprechend - signifikante Beziehungen nachweisen ließen. Damit benutzen wir kreuzweise das eine Verfahren als 'Kriterium' für das andere Verfahren, - obwohl wir von keinem der beiden die Gültigkeit kennen. Solches Vorgehen läßt sich allein für den Fall theoretisch rechtfertigen, daß diese Inter-Relationen Hypothesen bestätigen, die unabhängig von beiden Verfahren aufgestellt wurden : wir erhielten - auf der Grundlage beider Verfahren - Ergebnisse, die unseren (allgemeinen) Erwartungen über das Verhältnis von Index und Indiziertem entsprechen. Zwar wissen wir nicht, 'wie genau' das Q_s-sort-Verfahren mißt : doch können wir folgern, daß man etwa anhand jener acht items, die wir zur Messung von nAggression konstruierten, tatsächlich

eine 'gewisse Auskunft' über die Art, wie sich die Versuchsperson in bezug auf diese Variable selbst einschätzt, erhält.

Da aber keine Anhaltspunkte dafür gegeben sind, 'wie genau' diese Auskunft ist, können wir Ergebnisse, die sich auf einzelne Variablen beziehen, nicht verallgemeinern (s.u.). Wir wissen nicht, in welcher Hinsicht unsere Ergebnisse bei validierten Meßinstrumenten modifiziert würden. Ebenso läßt sich nicht entscheiden, welcherart Ergebnisse zu erwarten sind, wenn man für die Wunschprobe ein bereits bewährtes Auswertungssystem heranziehen könnte, nach welchem man etwa die Testprotokolle zusätzlich nach ihrer Ergiebigkeit und die einzelnen Antworten danach gewichtete, in welchem Grade sie inhaltlich auf die betreffende Variable verweisen.

Schließlich ist auf das Problem zu verweisen, daß beides, die Konstruktion der Q_s-sort-items wie die Signierung der WP-Antworten durch denselben Autor vorgenommen wurden, also durch dasselbe 'bias' bestimmt sind. Die Untersuchung kann nur Information über jene 'theoretische Realität' erbringen, welche durch die experimentellen Operationen 'realisiert' wurde : es ist nicht g e s i c h e r t , daß durch die verwendeten Operationen gerade die 'theoretische Realität' experimentell erfaßt wurde, die - unserer Intention nach - zu untersuchen war.

Ein weiteres Problem, das mit dem obigen unmittelbar zusammenhängt, besteht darin, daß wir wohl versuchten, unsere Grundannahmen experimentell zu 'r e a l i s i e r e n', aber nicht den Versuch machten, diesen Ansatz zu 'f a l s i f i z i e r e n'. Mit Popper ist der empirische Wert einer Theorie proportional dem Grad ihrer 'Falsifizierbarkeit' : der Grad ihrer 'Bewährung' (corroboration) ist um so größer, je mehr Versuche ihrer Falsifikation erfolglos waren.

Wir haben ein E r k u n d u n g s -Experiment durchgeführt. Die experimentelle Anordnung wurde ohne Theorie im strengen Sinne, ohne Hypothesen, die auf dem Wege der

Deduktion gewonnen worden wäre, vorgenommen. Unter relativ 'globalen' Gesichtspunkten wurde - ohne Validitätskontrolle und bei einer geringen Anzahl von Versuchspersonen - untersucht, o b sich 'irgendwelche' Beziehungen zwischen Test- und Kriteriumsverfahren aufweisen lassen, ein Vorgehen, das wir mit Wilkinson (1951) als 'blind analysis type of experiment' bezeichneten. Der Maßstab der 'Bewährung' kann – auf der Grundlage solchen Vorgehens - an das theoretische Konzept nicht angelegt werden. Es war von Anfang an nicht das Ziel dieser Untersuchung, Allgemeinaussagen herzuleiten, seien sie deduktiv (wie es der Begriff der Bewährung impliziert) oder induktiv gewonnen. Somit kommen wir zu dem Problem, welcherart Schlußfolgerungen überhaupt den vorliegenden experimentellen Daten zu entnehmen sind.

2. Folgerungen aus den experimentellen Ergebnissen

Es sei zunächst noch einmal abgegrenzt, welcherart Information sich diesem Experiment n i c h t , jedenfalls nicht gesichert entnehmen läßt.

Angesichts aller Einschränkungen, die wir im obigen Abschnitt noch einmal zusammengefaßt haben, können wir keinerlei Aussagen machen, die sich auf e i n z e l n e V a r i a b l e n beziehen. Es wurde nicht 'die' Aggressivität in ihren verschiedenen 'Erscheinungsmodi' (sprich 'Ebenen') gemessen, auch nicht 'Aggressivität' im Sinne jener Definition, an welcher die Konstruktion der Q_S-sort-items ausgerichtet war. Das bedeutet : wenn wir verschiedene Ergebnisse z.B. für nAggression und nDominance erhielten (etwa Werte, nach welchen hohe Wunsch-Dom-scores mit einem hohen Motivausdruck in der Wunschprobe (+Dom), hohe Wunsch-Agg-scores aber mit starker Ablehnung von nAgg (-Agg) verknüpft sind), so können wir nicht folgern, daß der Motivausdruck für diese Variablen sich a l l g e m e i n in dieser Hinsicht unterscheidet. Allerdings wird durch diese Ergebnisse nahegelegt, mit Hilfe validierter Meßinstrumente einen Entscheidungsversuch darüber

durchzuführen, ob die Test-scores dieser beiden Variablen (und entsprechend die der übrigen Variablen) tatsächlich in solch unterschiedlicher Weise zu interpretieren sind.

Für a l l e Variablen ließ sich - bei dieser experimentellen Anordnung - zeigen, daß die need-scores der Wunschprobe zumindest einer oder auch - bei verschiedenen Untergruppen - verschiedenen Ebenen zugeordnet werden können. Es ließ sich ferner zeigen, daß die Relationen zwischen Test-scores und Kriterium dann am engsten sind, wenn man gleichzeitig z w e i Ebenen, d.h. zwei Ebenen in ihrem Verhältnis zueinander als Kriterium benutzt. Betrachtete man dieses Verhältnis der Selbsteinschätzungen zweier Ebenen, so war eine inhaltliche, eine 'plausible' Interpretation der besonderen Eigenart der Testreaktion eher möglich als für den Fall, daß als 'Kriterium' allein die Daten e i n e r Ebene herangezogen wurden. Diese Tatsache, daß insbesondere Daten hinsichtlich der Beziehung zweier Ebenen (also etwa real-self - ideal-self) eine Interpretation des Testverhaltens ermöglichen, werteten wir dahingehend, daß erst eine Betrachtung des Ebenen- G e f ü g e s der Person das Verhältnis von 'Index und Indiziertem' erkennbar macht.

Allerdings handelt es sich hier um eine relativ unspezifische Schlußfolgerung. Denn es blieb - wieder für diese besondere experimentelle Anordnung - ungeklärt, durch welche Meßwerte das Verhältnis zweier Ebenen darzustellen sei : Ist schon die Tatsache einer Differenz zwischen zwei Ebenen als solche aufschlußreich, oder ist erst das zusätzliche Merkmal ausschlaggebend, daß die Vp durch eine (inter-individuell gesehen) extrem positive oder negative Selbsteinschätzung auf einer der beiden Ebenen charakterisiert ist? Es wurde insbesondere nicht untersucht, ob sich die Aussagemöglichkeiten eines Tests resp. der Wunschprobe dann sehr viel präziser erfassen lassen, wenn man das Verhältnis gleichzeitig aller d r e i Ebenen in Betracht zieht.

Eine weitere Frage blieb unbeantwortet : die Frage nach der Rolle jener Meßwerte, die man in der Literatur als

Indiz einer allgemeinen 'Angepaßtheit' oder 'Stabilität' aufgefaßt hat, nämlich das Verhältnis zwischen zwei 'Ebenen insgesamt gesehen'. Wir konnten nicht entscheiden, ob eine hohe durchgängige Übereinstimmung etwa zwischen Wunsch- und ideal-Ebene sich in irgendeiner Weise auf die Testreaktion auswirkt und zudem in a n d e r e r Weise als etwa eine hohe D i f f e r e n z zwischen beiden Ebenen. Zwar erhielten wir für derart verschiedene Gruppen unterschiedliche Korrelationen zwischen Test und Kriterium, doch blieb ungeklärt, welche statistische Signifikanz ihnen zuzuschreiben ist. Zu bemerken ist hier, daß das allgemeine Kriterium des Verhältnisses zwischen 'zwei Ebenen insgesamt gesehen' allein auf den Motivausdruck e i n z e l n e r Variablen bezogen wurde : wir haben nicht geprüft, ob dieses allgemeine Kriterium in Zusammenhang gebracht werden kann mit ebenfalls allgemeineren Merkmalen der Testprotokolle, etwa ihrer Ergiebigkeit, Länge, ihrer sprachlichen Differenziertheit etc. Alle übrigen formalen oder inhaltlichen Merkmale der Testprotokolle - außer jenen, die sprachlich einzelnen need-Kategorien zugeordnet werden konnten - blieben in dieser Untersuchung unberücksichtigt.

Für jene Merkmale der Testprotokolle jedoch, die wir anhand unseres need-Schemas signierten, also quantifizierten, ließen sich systematische Relationen zwischen Test und Kriterium nachweisen : die Test-scores für einen einzelnen need konnten - statistisch gesichert - den verschiedenen 'Erscheinungsmodi' eben dieses 'need's' zugeordnet werden. Wir zogen daraus den Schluß, daß derartige Testreaktionen, die als Ausdruck von 'needs' oder 'Eigenschaften' verstanden werden können, nur dann sich in ihren Bedingungen und damit in ihrer B e d e u t u n g erfassen lassen, wenn man die betreffende Eigenschaft auf allen ihren 'Erscheinungsebenen' analysiert. Diese Schlußfolgerung ist auf ihre Allgemeingültigkeit hin zu überprüfen.

Aus unserem Experiment läßt sich als H y p o t h e s e ableiten, daß die Wunschprobe insbesondere solche psycho-

logischen Gegebenheiten mißt, welche sich nur in Begriffen des Ebenen-Konzeptes hinreichend erfassen lassen. Diese Hypothese wäre einer umfangreichen Untersuchung zu unterziehen : die experimentellen Ergebnisse zeigen, daß eine solche Analyse notwendig ist, soll die Gültigkeit des Tests bestimmt werden können.

Es kann jedoch nicht als bewiesen gelten, daß bei jedem oder jedenfalls bei jedem projektiven Test der Motivausdruck n u r in den Kategorien des Ebenen-Konzeptes auf seine diagnostische Relevanz hin analysiert werden kann. Die Schlußfolgerungen, die sich dem Experiment entnehmen lassen, sind auf die W u n s c h p r o b e begrenzt.

Gleichwohl nehmen wir an, daß bei a l l e n projektiven Verfahren jene Testreaktionen, die sich n e e d - Kategorien zuordnen lassen, nicht ohne das Konzept verschiedener 'Erscheinungsmodi' eines needs analysiert werden können. Projektive Tests sind - so weit lassen sie sich einem gemeinsamen Begriff unterordnen - insbesondere dadurch charakterisiert, daß dem Probanden m e h r - d e u t i g e s stimulus-Material dargeboten wird : dieses läßt eine Vielfalt von Assoziationen zu und liefert damit eine Vorbedingung für d e f e n s e -Verhalten gegen die eine oder andere Assoziation. Damit aber sind die Voraussetzungen dafür gegeben, daß die Reaktion auf dieses stimulus-Material bewußt oder unbewußt g e s t e u - e r t werden kann. Diese Steuerung bezieht sich auf die Wahrnehmung, die Selektion der stimuli, wie auf die Beantwortung des Tests. Wahrnehmungshypothesen oder 'expectancies' können u n m i t t e l b a r e r Ausdruck einer bestimmten Bedürfnisstruktur sein, die Reaktion kann 'ungehindert' der unmittelbaren Bedürfnisbefriedigung dienen. Beide, Wanrnehmung wie Reaktion,aber können sozusagen kognitiv 'g e f i l t e r t' werden. Daraus folgt, daß die Beantwortung eines projektiven Tests notwendig in den zwei Dimensionen von 'Antrieb' und 'Steuerung' oder auch 'need' und 'regulation' zu sehen ist. Die Art der Steuerung der Testreaktion ist - so nehmen wir ferner an - ihrerseits nur in den Begriffen von 'need' und 'regulation' erfaßbar, - 'need' und 'regulation' hier nicht verstanden

Indiz einer allgemeinen 'Angepaßtheit' oder 'Stabilität' aufgefaßt hat, nämlich das Verhältnis zwischen zwei 'Ebenen insgesamt gesehen'. Wir konnten nicht entscheiden, ob eine hohe durchgängige Übereinstimmung etwa zwischen Wunsch- und ideal-Ebene sich in irgendeiner Weise auf die Testreaktion auswirkt und zudem in a n d e r e r Weise als etwa eine hohe D i f f e r e n z zwischen beiden Ebenen. Zwar erhielten wir für derart verschiedene Gruppen unterschiedliche Korrelationen zwischen Test und Kriterium, doch blieb ungeklärt, welche statistische Signifikanz ihnen zuzuschreiben ist. Zu bemerken ist hier, daß das allgemeine Kriterium des Verhältnisses zwischen 'zwei Ebenen insgesamt gesehen' allein auf den Motivausdruck e i n z e l n e r Variablen bezogen wurde : wir haben nicht geprüft, ob dieses allgemeine Kriterium in Zusammenhang gebracht werden kann mit ebenfalls allgemeineren Merkmalen der Testprotokolle, etwa ihrer Ergiebigkeit, Länge, ihrer sprachlichen Differenziertheit etc. Alle übrigen formalen oder inhaltlichen Merkmale der Testprotokolle - außer jenen, die sprachlich einzelnen need-Kategorien zugeordnet werden konnten - blieben in dieser Untersuchung unberücksichtigt.

Für jene Merkmale der Testprotokolle jedoch, die wir anhand unseres need-Schemas signierten, also quantifizierten, ließen sich systematische Relationen zwischen Test und Kriterium nachweisen : die Test-scores für einen einzelnen need konnten - statistisch gesichert - den verschiedenen 'Erscheinungsmodi' eben dieses 'need's' zugeordnet werden. Wir zogen daraus den Schluß, daß derartige Testreaktionen, die als Ausdruck von 'needs' oder 'Eigenschaften' verstanden werden können, nur dann sich in ihren Bedingungen und damit in ihrer B e d e u t u n g erfassen lassen, wenn man die betreffende Eigenschaft auf allen ihren 'Erscheinungsebenen' analysiert. Diese Schlußfolgerung ist auf ihre Allgemeingültigkeit hin zu überprüfen.

Aus unserem Experiment läßt sich als H y p o t h e s e ableiten, daß die Wunschprobe insbesondere solche psycho-

logischen Gegebenheiten mißt, welche sich nur in Begriffen des Ebenen-Konzeptes hinreichend erfassen lassen. Diese Hypothese wäre einer umfangreichen Untersuchung zu unterziehen : die experimentellen Ergebnisse zeigen, daß eine solche Analyse notwendig ist, soll die Gültigkeit des Tests bestimmt werden können.

Es kann jedoch nicht als bewiesen gelten, daß bei jedem oder jedenfalls bei jedem projektiven Test der Motivausdruck n u r in den Kategorien des Ebenen-Konzeptes auf seine diagnostische Relevanz hin analysiert werden kann. Die Schlußfolgerungen, die sich dem Experiment entnehmen lassen, sind auf die W u n s c h p r o b e begrenzt.

Gleichwohl nehmen wir an, daß bei a l l e n projektiven Verfahren jene Testreaktionen, die sich n e e d - Kategorien zuordnen lassen, nicht ohne das Konzept verschiedener 'Erscheinungsmodi' eines needs analysiert werden können. Projektive Tests sind - so weit lassen sie sich einem gemeinsamen Begriff unterordnen - insbesondere dadurch charakterisiert, daß dem Probanden m e h r - d e u t i g e s stimulus-Material dargeboten wird : dieses läßt eine Vielfalt von Assoziationen zu und liefert damit eine Vorbedingung für d e f e n s e -Verhalten gegen die eine oder andere Assoziation. Damit aber sind die Voraussetzungen dafür gegeben, daß die Reaktion auf dieses stimulus-Material bewußt oder unbewußt g e s t e u e r t werden kann. Diese Steuerung bezieht sich auf die Wahrnehmung, die Selektion der stimuli, wie auf die Beantwortung des Tests. Wahrnehmungshypothesen oder 'expectancies' können u n m i t t e l b a r e r Ausdruck einer bestimmten Bedürfnisstruktur sein, die Reaktion kann 'ungehindert' der unmittelbaren Bedürfnisbefriedigung dienen. Beide, Wahrnehmung wie Reaktion,aber können sozusagen kognitiv 'g e f i l t e r t' werden. Daraus folgt, daß die Beantwortung eines projektiven Tests notwendig in den zwei Dimensionen von 'Antrieb' und 'Steuerung' oder auch 'need' und 'regulation' zu sehen ist. Die Art der Steuerung der Testreaktion ist - so nehmen wir ferner an - ihrerseits nur in den Begriffen von 'need' und 'regulation' erfaßbar, - 'need' und 'regulation' hier nicht verstanden

als a k t u e l l e Determinanten des T e s t verhaltens, sondern gefaßt als allgemeine Dimensionen der Persönlichkeit. Diese Dimensionen beschrieben wir als die 'Ebenen' von Bedürfnis, Anspruch und tatsächlichem Verhalten. Die Steuerung einer Aggressions-Reaktion (inhibition) oder auch ihr unmittelbarer Ausdruck (expression) sind in ihren Bedingungen nur zu verstehen, wenn man 'Aggression' in ihren verschiedenen 'Erscheinungsmodi' erfaßt und das heißt : die Ebenen von Bedürfnis, Anspruch und Verhalten in ihrem gegenseitigen G e f ü g e analysiert.

Es ist dies die theoretische Grundannahme, welche zu dem vorliegenden Experiment führte; sie ist indessen aus diesem Experiment nicht als erwiesen abzuleiten. Um die obige These zu verifizieren, wären Reaktionen auf mehrdeutiges stimulus-Material in ihrem 'Ablauf', in ihrem 'Prozeß' eigens zu analysieren. Allerdings erscheint es fraglich, ob in diesem Bereich 'hypothetical constructs' vollständig in meßbare 'intervenierende Variablen' überführt werden können.

Auf der Grundlage der vorliegenden experimentellen Daten können wir nur folgern : das Ebenen-Konzept ist dazu geeignet, Testreaktionen auf die Wunschprobe zu analysieren. Es liefert Anhaltspunkte für die Zuordnung von Test-Index und Kriterium; es führt - bei diesem Test - zu Hinweisen auf eine differentielle Validität.

Auch für diese Frage einer differentiellen Validität aber sind dem Experiment keine gesicherten Ergebnisse zu entnehmen. Es ließ sich zeigen, daß die Beziehung zwischen Test und Kriterium für einzelne U n t e r gruppen, die anhand eines bestimmten Merkmals zusammengestellt wurden, e n g e r (nämlich korrelativ) ist als für die gesamte Versuchsgruppe. Derartige Merkmale, welche eine Untergruppe konstituieren, wurden insbesondere auf der Grundlage der Daten e i n z e l n e r Variablen gefunden : es ist die einheitliche (extreme) Selbsteinschätzung e i n e r Variablen auf e i n e r Ebene, welche eine Untergruppe konstituiert, für die sich eine gleichartige Reaktion auf die Wunschprobe nachweisen ließ.

Ungeklärt blieb - wie schon einmal erwähnt - , ob auch der Beziehung zwischen zwei 'Ebenen insgesamt' die Rolle von Trennkriterien zukommt. Es könnten unsere Ergebnisse aber darauf verweisen, daß ein solches Merkmal ('Ebenen insgesamt') nicht eine einheitliche Bedingung für den Motivausdruck a l l e r Variablen beschreibt. Sofern man diesen Ergebnissen überhaupt etwas entnehmen kann, zeigen sie an, daß auch bei einer Untergruppe, die nach einem derartigen 'allgemeinen Merkmal' gebildet wird, der Ausdruck verschiedener needs nicht ein und derselben Ebene zugeordnet werden kann. Das würde bedeuten : das Testprotokoll e i n e s Probanden ist - je nach need-Kategorie - auf v e r s c h i e d e n e n Ebenen auszuwerten.

Auch diese Frage läßt sich anhand unseres Datenmaterials nicht klären. Indessen ist sie - von unserem Grundansatz her - t h e o r e t i s c h eindeutig zu beantworten. Wenn es richtig ist, daß Testreaktionen mehr oder weniger 'gesteuert' werden und wenn es zutrifft, daß dieser Grad der Steuerung abhängig ist von der Art des 'Ebenen-Gefüges', dann ist für jeden need ein eigenes Bedingungsgeflecht zu erwarten. Bei ein und derselben Vp unterscheidet sich nAggression von nNurturance und dieses von nHarmavoidance hinsichtlich der Relation von Bedürfnis, Anspruch und Verhalten : für jede Variable sind andere Bedingungen für die Testreaktion gegeben, und gleichförmige Bedingungen werden wieder bei verschiedenen Variablen von unterschiedlicher Bedeutung für die Art des Motivausdrucks sein. Eine solche für verschiedene needs unterschiedliche Funktionsweise eines Tests aber führt zu Konsequenzen, die für die Diagnostik höchst problematisch sind. Sie werden weiter unten noch einmal untersucht.

Anhaltspunkte für eine differentielle Validität liefern die experimentellen Daten nur in sehr unzureichender Weise, und zwar vor allem deshalb, weil - erklärtermaßen - die Versuchspersonen eine relativ homogene Gruppe bilden. Es steht aus, Gruppen verschiedenen Geschlechts, Alters, Sprachschatzes, verschiedener sozialer Herkunft, Intelligenz und Ausbildung etc. zu untersuchen. Es ist zu erwarten, daß andere Dimensionen, wie Interessen, Einstellun-

gen, Gewohnheiten, die Art der Testreaktion mit bestimmen. Wahrscheinlich ist aber auch, daß Interessen und Einstellungen einen wesentlichen Bestandteil dessen ausmachen, was der Wunschprobe an diagnostisch Relevantem entnommen werden kann. Diese Frage wollen wir ebenfalls abschließend noch einmal aufgreifen.

Anknüpfend an die allgemeinen Leitgesichtspunkte, die wir dem Experiment voranstellten, fassen wir zusammen :

Unserem Material nach funktioniert die Wunschprobe bei verschiedenen Variablen, wie bei verschiedenen Gruppen auf verschiedenen 'Ebenen'. Auch jene Kriterien ließen sich anhand der Begriffe des Ebenen-Konzepts bestimmen, mit welchen 'population differences in construct validity' beschrieben werden können. Ob die Wunschprobe bei einzelnen Variablen für alle Versuchspersonen gleichartig auf nur e i n e r 'Funktionsebene' zu interpretieren ist (z.B. bei nDominance, nHarmavoidance) blieb ungeklärt.

3. Probleme und Konsequenzen einer differentiellen Validität

Unseren Ergebnissen zufolge ist das Testprotokoll e i n e s Probanden - je nach need-Kategorie - auf v e r s c h i e d e n e n Ebenen zu interpretieren, - ein Tatbestand, der aufgrund unseres theoretischen Ansatzes zu erwarten war. Dem Datenmaterial nach sind auch - diese Konsequenz ist allerdings statistisch nicht genügend geklärt - die need-scores e i n e r V a r i a b l e n bei verschiedenen Versuchspersonen v e r s c h i e d e n e n Ebenen zuzuordnen. Zumindest aber ließ sich zeigen, daß d i e s e l b e Variable bei verschiedenen Versuchspersonen auf der Grundlage verschiedener inter-individueller N o r m e n zu interpretieren ist : nur für einzelne Untergruppen, die anhand eines bestimmten Merkmals zusammengestellt worden waren, konnte eine k o r r e l a t i v e Beziehung zwischen Test-Index und Kriterium nachgewiesen werden; während für die gesamte Gruppe die inter-individuelle Differenzierung sehr viel 'gröber', nämlich nur mittels Extremgruppenmethode aufweisbar ist.

Der Begriff 'differentielle Validität' bezeichnet den Tatbestand, daß derselbe Test nicht notwendig für a l l e Probanden valide ist und daß er für verschiedene Probanden in 'unterschiedlichem Grade' oder auch auf 'unterschiedliche Weise' valide ist, d.h. daß der Test nach verschiedenen 'Normen' oder auch nach verschiedenen 'Regeln' ausgewertet werden muß.

Unseren Ergebnissen nach ist es wahrscheinlich, daß die Wunschprobe bei verschiedenen Gruppen nach verschiedenen Normen zu interpretieren ist oder auch, daß nur für bestimmte Untergruppen sich überhaupt Normen aufstellen lassen, - Normen, welche ein K o n t i n u u m von interindividuell unterschiedlichen Ausprägungsgraden eines Merkmals beschreiben.

Sollten für verschiedene Probanden bzw. Gruppen von Probanden nur derart unterschiedliche Test n o r m e n zu bestimmen sein, wäre die Aufgabe einer Testvalidierung noch vergleichsweise einfach. Ist stattdessen damit zu rechnen, daß das 'Indizierte' nicht allein seiner quantitativen Ausprägung, sondern auch seiner qualitativen Eigenart nach bei verschiedenen Probanden variiert, so ist das Vorhaben einer Gültigkeitsbestimmung ungleich problematischer. Der letztere Fall ist gegeben, wenn die Motivscores e i n e r Variablen bei verschiedenen Versuchspersonen auf v e r s c h i e d e n e n E b e n e n zu interpretieren sind.

Wir wiesen oben schon darauf hin, daß das Datenmaterial nicht klar erkennen läßt, ob dieser Fall einer unterschiedlichen Gültigkeit auf die Wunschprobe zutrifft. Indessen ist unserem theoretischen Ansatz nach zu erwarten, daß für die Wunschprobe eine derartige Form der differentiellen Validität gegeben ist. Wir nehmen an, daß der Motivausdruck im wesentlichen durch die Art des Ebenen-G e f ü g e s bestimmt wird : und dieses Ebenen-Gefüge ist - selbstverständlich - auch für e i n e Variable bei verschiedenen Versuchspersonen n i c h t dasselbe.

Der Motivausdruck einer bestimmten Variablen würde sich d a n n für alle Vpn ein und derselben Ebene zuordnen lassen, wenn folgende Voraussetzung erfüllt wäre : wenn

die inter-individuellen Unterschiede hinsichtlich des Ebenen-Gefüges sich auf einem g e m e i n s a m e n K o n t i n u u m anordnen ließen. Parallel zu diesen Unterschieden würde dann - wie sogleich zu erläutern ist - der Motivausdruck auf e i n e r Ebene variieren oder auch Grad und Richtung der Relation zwischen z w e i Ebenen indizieren.

Unserem Material nach - wir nehmen hier noch einmal kurz auf die experimentellen Daten bezug (vgl.Tab. VI, VII, S.235, 237) - ist dieser Fall nicht auszuschließen. Für drei Variablen (nDom, nNur, avoid) ließ sich auf der Grundlage der g e s a m t e n Versuchsgruppe für den Vergleich von Wunschprobe und Kriterium eine Extremgruppen-Differenzierung feststellen, die - weil kreuzweise nachgewiesen - statistisch hinreichend gesichert ist. Für eine Variable (nDom) ergab sich sogar eine k o r r e l a t i v e Beziehung zwischen Test und Kriterium : Ebenso wie nicht bekannt ist, ob diese Korrelation zufallsbedingt ist, kann nicht ausgeschlossen werden, daß eine solche korrelative Beziehung auch für a n d e r e Variablen nachweisbar wäre, und zwar für den Fall, daß 'gültige' Meßinstrumente benutzt würden.

Als Kriterium fungierten hier die Differenzen zwischen zwei Ebenen : ist also - so wäre hier zu fragen - diese Differenz zwischen zwei Ebenen für alle Versuchspersonen gleichermaßen a l l e i n ausschlaggebend für die Testreaktion? Bilden etwa für die Variable nDominance die Differenzen zwischen ideal- und Wunsch-Ebene ein inter-individuell einheitliches Kontinuum, dem der Motivausdruck in der Wunschprobe analog variierte? Der Motiv-score wäre dann bei allen Versuchspersonen Indiz für den Grad der Übereinstimmung oder Differenz gerade dieser zwei Ebenen. Sind es demgegenüber für die Variable nNurturance die Differenzen zwischen real- und ideal-Ebene, für nHarmavoidance die Differenzen zwischen real- und Wunsch-Ebene, welche für alle Vpn gleichartig durch die betreffenden need-scores in der WP angezeigt würden (vgl.Tab. VI und VII)? Ist also für diese Variablen die jeweils dritte Ebene (bei nDom die Ebene des Verhaltens, bei nNur die Wunsch-Ebene) irrelevant für die Art des Motivausdrucks?

Die Ergebnisse können in der Tat dahingehend verstanden werden, daß das 'Indizierte' inter-individuell gesehen wohl quantitativ, nicht jedoch qualitativ variiert. Das heißt : die Motiv-scores wären hier nicht bei verschiedenen Probanden 'auf verschiedenen Ebenen' zu interpretieren, also n i c h t 'nach verschiedenen Regeln' auszuwerten. Die Test-scores wären vielleicht - bei verschiedenen Gruppen von Vpn - auf verschiedene N o r m e n zu beziehen, was dann wahrscheinlich ist, wenn dieselbe Relation bei einigen Untergruppen - im Gegensatz zur Gesamtgruppe - als korrelative Beziehung feststellbar ist.

Gleichwohl läßt sich - allein auf der Basis unserer Daten - nicht entscheiden, ob diese Relationen, nach welchen der Test bei allen Versuchspersonen dasselbe indizierte, verallgemeinert werden können. Es wurde nur eine relativ homogene Versuchsgruppe herangezogen : daher bleibt noch zu untersuchen, welchem Kriterium die Wunschprobe bei anderen Gruppen zuzuordnen ist. Durch diese Diskussion aber wurde deutlich, daß hier ein entscheidendes Problem vorliegt; die Frage nach den Bedingungen der differentiellen Validität bedarf in der weiteren Forschung noch einer sehr differenzierten Analyse : Sind es nur quantitative, oder sind es qualitative Unterschiede des Motivausdrucks, welche die differentielle Validität ausmachen?

Die Dringlichkeit dieser Frage wird deutlich, wenn man die Konsequenzen für den Fall bedenkt, daß nicht - wie die obigen Werte nahelegen könnten - a l l e Versuchspersonen sich hinsichtlich ihres 'Ebenen-Gefüges' einem g e m e i n s a m e n Kontinuum zuordnen lassen, sondern wenn - was wiederum durch andere Werte indiziert werden könnte (vgl. Tab. X, s.S. 262) - die Test-scores verschiedener Vpn 'auf verschiedenen Ebenen' zu interpretieren wären.

Im ersteren Falle wäre für jede Variable bekannt, w a s sie indiziert : es wäre etwa festgestellt worden, daß die Test-scores für nDominance bei allen Vpn ein verläßliches Indiz für den Grad wie die Richtung der Differenz zwischen Bedürfnis und Verhaltens-Anspruch darstellen. Es bliebe - als Folge der differentiellen Validität des Tests - allein festzustellen, welche inter-individuellen Normen als Vergleichsmaßstab heranzuziehen sind, um der Versuchsperson einen relativen Ort innerhalb einer bestimmten Bezugs-

(bzw. Norm-)Gruppe zuzuweisen.

Im zweiten Falle jedoch wäre zunächst zu klären, für welcherart psychologischen Phänomens die Wunschprobe für die besondere Versuchsperson 'empfindlich' war; in den Begriffen des Ebenen-Konzeptes gesprochen : auf welcher Ebene sie gerade bei diesem Probanden 'funktionierte'. Und diese Information wäre ihrerseits wieder für jede einzelne Variable getrennt zu gewinnen : denn es ist nicht - wir haben es oben erörtert - damit zu rechnen, daß der Test bei e i n e r Versuchsperson für die verschiedenen Variablen durchgehend dieselbe Ebene indizierte. Sodann wäre festzustellen, welche quantitativen Normen heranzuziehen sind, damit der Test-scores etwa für nAggression, der bei diesem Probanden vielleicht - wie festgestellt worden wäre - die Antriebsseite indizierte, überhaupt diagnostisch und das heißt hier differentiell-diagnostisch verwertet werden kann. Dies aber wäre zweifellos ein zu aufwendiges Unternehmen, um für die Diagnostik noch praktikabel zu sein.

Da wir nicht ausschließen können, daß dieser Fall einer differentiellen Validität auf die Wunschprobe zutrifft, wollen wir dessen Konsequenzen durchdenken : an diesen Konsequenzen kann, wie wir schon vermerkten, deutlich gemacht werden, weshalb eingehende Analysen zur Validität der Wunschprobe (als Beispiel eines projektiven Verfahrens) dringend notwendig sind, soll nicht ein solcher Test zu höchst irreführenden diagnostischen Schlüssen verleiten.

Die einfachste, wenngleich am wenigsten befriedigende Konsequenz ist die, daß man auf eine solche (differentielle) Validierung einzelner Meßgrößen (Variablen) für einen Test wie den der Wunschprobe v e r z i c h t e t e .

Als Alternative wäre in diesem Falle geltend zu machen, daß es größeren Erfolg versprechen könnte, für die verschiedenen need-Kategorien jeweils einen e i g e n e n Test zu konstruieren.

Ein solcher Test könnte - vielleicht - gerade jene Voraussetzungen erfüllen, welche die Wunschprobe nicht erfüllt. Das Problem der Gültigkeitsbestimmung besteht bei der Wunschprobe darin, daß das zu messende Kriterium zu komplex ist, als daß es in seinen verschiedenen Aspekten an-

hand der Wunschproben-Protokolle signiert werden könnte. Genauer : Motivmessung setzt voraus, daß man das Motiv im Bezugssystem aller Ebenen analysieren, daß man Antrieb u n d Steuerung in ihrem Wechselspiel erfassen kann. Ein need-score in der Wunschprobe zeigt aber a l s s o l c h e r nicht an, ob er auf ein fehlendes Bedürfnis, ein schwach ausgeprägtes Verhalten, einen geringen Anspruch oder aber auf ein starkes, jedoch defensiv abgewehrtes Bedürfnis hinweist. Er kann - wie wir oben aufzeigten - nur interpretiert werden, wenn entweder nachgewiesen ist, daß die WP bei dieser need-Kategorie durchgängig (also bei allen Vpn) beispielsweise die Ich-Ideal-Ebene anzeigte oder wenn zuvor (etwa durch Vortests) festgestellt ist, w a s die Wunschprobe in diesem besonderen Falle indiziert.

Ein Test jedoch, der allein zur Messung einer einzigen Variablen konstruiert worden wäre, könnte es vielleicht leisten, diese verschiedenen Dimensionen g e t r e n n t zu messen, 'need' und 'regulation' e i n z e l n zu bestimmen : aus der relativen Stärke dieser zwei Tendenzen könnte dann - vielleicht - o h n e Vortests direkt auf das Bedürfnis in seinem Verhältnis zum 'Verhalten' und 'Verhaltensanspruch' geschlossen werden. Derartige Verfahren hat Farberow (1959) vorgeschlagen, - ohne allerdings genügend konkrete Anweisungen für ihre Realisation zu geben. Farberow schlug vor, für eine einzelne need-Kategorie ein nach dem Grade der stimulus-relevance abgestuftes stimulus-Material darzubieten : 'gemessen' wird, welche Grade von stimulus-relevance nötig sind, um den A u s d r u c k des betreffenden Motivs oder aber dessen V e r m e i d u n g zu bewirken (vgl. S. 80f). Allerdings ist auch hier - diese Konsequenz wurde von Farberow nicht gesehen - mit einer d i f f e r e n t i e l l e n Validität zu rechnen, da nicht die stimulus-relevance als solche, sondern die '(differentiell) erlebte stimulus-Qualität' die Reaktion bestimmt.

Auf eine weitere Möglichkeit für die Messung einer einzelnen Motivvariablen verweisen die Arbeiten von Lesser :

was er für die Variable nAggression aufzeigte, ist möglicherweise auch auf andere Variablen zu übertragen : daß nämlich 'need' und 'regulation' b e i d e und das heißt 'thematisch' zum Ausdruck kommen und so in ihrer relativen Stärke demselben Test entnommen werden können.

Ein anderer Ausweg aus solcherart Problemen der Motivmessung bestände - theoretisch - darin, Verfahren zu konstruieren, welche von diesen zwei Tendenzen, von 'need' und 'regulation', u n a b h ä n g i g wären. Es handelte sich nicht um projektive Verfahren und gewiß nicht um mehrdeutiges stimulus-Material. Verfahren dieser Art könnten nur auf physiologischen Methoden beruhen, weil anders sie wohl kaum dem 'Willen', also der 'Steuerung' entzogen werden könnten. Man würde etwa die Geschwindigkeit, mit der ein bedingter Vermeidensreflex gelernt wird, als Indikator verwenden.

Hier stoßen wir jedoch auf ein Problem, das eine ausführliche Diskussion verlangte und eine eigene Arbeit lohnte : das Problem des Verhältnisses von Diagnostik und Behaviorismus und das vergleichbare Problem des Verhältnisses von Psychoanalyse und einer Verhaltenstheorie etwa der Eysenck'schen Prägung. Wir können es in diesem Rahmen nicht erörtern. Wir verweisen nur darauf, was wir an anderer Stelle darlegten, daß Persönlichkeitsdiagnose nicht ohne Begriffe des s o z i a l e n Verhaltens möglich ist; wir fügen an, was ebenfalls ausführlich zu diskutieren wäre, daß die Analyse 'sozialen Verhaltens' ihrerseits soziologische und - was am wenigsten sich mit dem Behaviorismus vereinbaren ließe - historische Kategorien voraussetzt. Eine Motivmessung auf physiologischer oder rein verhaltenstheoretischer Grundlage halten wir nicht für möglich, jedenfalls nicht eine Messung jener Motive, die für eine Persönlichkeitsdiagnostik von Interesse sind. Welche Bedeutung kommt also dén p r o j e k t i v e n Methoden zu?

Wir kehren zu unserer Ausgangsfrage zurück : Welche Konsequenzen sind zu ziehen, wenn die Gültigkeit eines - bereits in der diagnostischen Praxis verwendeten - projektiven Verfahrens nur auf einem derart komplizierten Wege

bestimmt werden kann, wie er unter den Bedingungen einer 'differentiellen Validität' einzuschlagen wäre.

Eine Gegenfrage sei zunächst geprüft : ist es in der diagnostischen Praxis überhaupt von Interesse, derart verschiedene need-Variablen zu messen? Wird für eine Persönlichkeitsdiagnose überhaupt eine Auskunft über alle nur denkbaren Verhaltensvariablen benötigt? Sind es nicht besondere Fälle, in denen hier Information über nAggression, dort über nDominance und wieder bei einem anderen Probanden Auskunft über nHarmavoidance benötigt wird? Das heißt : sollten nicht für derart spezifische Merkmale jeweils einzelne Tests ausgewählt werden können, während man von einem projektiven Test eine solche spezielle Information gar nicht erwartet? Oder trifft es - auch für die projektiven Methoden - zu, was Hörmann (1964b) formulierte, daß die Praxis lernen müsse, speziellere Fragen an die Diagnostik zu stellen?

Diese Frage spiegelt die Entwicklung innerhalb der psychologischen Methodologie, die Entwicklung eines zunehmend kritischen Begriffsapparats, zunehmend verfeinerter Methoden in der empirischen Erfolgskontrolle, eines zunehmend verschärften Anspruchs an die Exaktheit der Messungen wie der begrifflichen Definition wider. Eine solche Entwicklung macht - auch diese Arbeit hat sich in diesen Rahmen gestellt - keineswegs halt vor den projektiven Methoden. Wenn diese Methoden es aber nicht leisten können, derart spezielle Fragen exakt genug zu beantworten - was indessen bis heute nicht entschieden ist - ,welche Funktion wäre dann dem herkömmlichen projektiven Test zuzusprechen?

Hier ist nun eine Überlegung anzustellen, die in der gesamten Arbeit bisher ausgeklammert blieb. Die Auswertung der Wunschprobe (als eines projektiven Tests) nach den Kategorien einzelner isolierter needs liefert nur einen - und vielleicht einen geringen - Ausschnitt der Aussagemöglichkeiten des Verfahrens. Bei aller wissenschaftlichen Insistenz darauf, daß 'gültige' Aussagen nur gewonnen werden können, wenn man sie auf den Vergleichsmaßstab m e ß b a r e r inter-individueller Differenzen bezieht,

ist doch nicht zu leugnen, daß etwa die Wunschprobe oder auch der TAT I d i o g r a p h i s c h e s hervorbringen, das - jedenfalls für den Diagnostiker - von nicht zu unterschätzender Bedeutung ist. Hinweise auf Idiographisches können nicht validiert werden, da sie - ex definitione - allein das Individuum in seiner Besonderheit betreffen; sie lassen sich nicht einem inter-individuellen Kontinuum zuordnen, so daß differentiell-psychologische Vergleiche möglich und durch den Maßstab einer Eichpopulation objektivierte Aussagen ableitbar wären. Ein derart unkontrollierter, d.h. nicht auf seine Verläßlichkeit geprüfter Umgang mit projektiven Verfahren aber gab gerade den Anlaß zu der Frage, welche Merkmale eines Testprotokolls als allgemeine S y m p t o m e zu bewerten sind, welche Merkmale in eine r e g e l h a f t e Beziehung zu einer psychologischen Gegebenheit gebracht werden können. Diese mit der Entwicklung der Forschung notwendig gewordene Frage nach der Validität auch der projektiven Tests hebt indessen nicht jene weiteren Aussagemöglichkeiten dieser Verfahren auf, die als idiographisch zu bezeichnen sind und dem Prinzip der Validität geradezu widersprechen.

Ein diagnostisches Verfahren m u ß nicht ein M e ß instrument sein, wenn auch dies - zu recht - das Ziel jeder wissenschaftlichen Bemühung um psychologische Tests ist. Ein Verfahren ist auch dann ein diagnostisches Hilfsmittel, wenn es nur als 'Auslöser' für ein diagnostisches Gespräch, für die Exploration, dient, wenn es - untheoretisch gesprochen - Hemmungen der Selbstdarstellung löst, wenn es etwa eine Art 'Medium' f ü r die Selbstdarstellung bietet, welche - im Einzelfall - in den gewohnten Bahnen der sprachlichen Kommunikation - aus welchem Grund auch immer - vielleicht nicht gelingen mag. Diagnostisch wertvoll ist ein Verfahren auch dann, wenn es besondere individuelle 'Themen', Problemzusammenhänge, subjektive Erfahrungen oder Lebensdaten hervorbringt : es hat dann eine hermeneutische Funktion und nicht die der quantitativen generalisierenden Messung.

Läßt sich eine solche hermeneutische Funktion - für den Fall, daß eine Gültigkeitsbestimmung undurchführbar wäre - angesichts der geltenden Maßstäbe wissenschaftlicher Psychologie rechtfertigen? Wir glauben, ja.

Persönlichkeitsdiagnose läßt sich nie - auch bei strengster Orientierung am positivistischen Wissenschaftsbegriff - reduzieren auf einfache Markierungspunkte innerhalb eines festgelegten Koordinatenkreuzes. Allein eine Eignungsuntersuchung im engsten Sinne kann auf einzelne Meßwerte gegründet werden, welche sich - unter Voraussetzung sozusagen eines 'Minimal-Kriteriums - auf die entsprechenden Meßwerte der Eichstichprobe beziehen lassen. Eine Persönlichkeitsdiagnose wird es immer notwendig machen, den Menschen nicht allein als Schnittpunkt verschiedener Merkmalsklassen zu sehen; ihre Aufgabe wird es - bei aller notwendigen Bemühung um objektivierte Aussagen - sein, den Menschen in seinem individuellen Lebenszusammenhang zu 'verstehen'.

Entfernen wir uns bedenklich von jenen Argumenten, die wir einleitend zur Begründung der vorliegenden experimentellen Untersuchung anführten? Wir glauben, nein. Aufgabe der Wissenschaft kann es allein sein, 'gültige', generalisierbare Aussagen zu gewinnen. Sie kann jedoch nicht jene Fragen, die sich dem methodischen Zugriff entziehen, aber gleichwohl den Gegenstand entscheidend mitbestimmen, schlechthin eliminieren : wenn sie diese Frage schon nicht beantworten kann, so hat sie gerade am Maßstab dieser ungelösten Fragen ihre eigenen Grenzen zu bestimmen.

Persönlichkeitsdiagnostik ist mit rein wissenschaftlichen Mitteln nicht möglich; sie muß dennoch betrieben werden. Diese Kluft ist allein durch den Diagnostiker zu überbrücken, - jedem kritischen Einwand zum Trotz - allein durch seine 'Erfahrung', seine 'Intuition', sein 'Sensorium', durch die Sensibilität seiner g e i s t i g e n Mittel. Die Alternative wäre der Verzicht auf den Diagnostiker und der Verlaß auf den Computer. Der Diagnostiker kann und darf selbstverständlich nicht einen Bestand an methodisch gesicherten, d.n. 'gültigen' Instrumenten ersetzen; aber :

die validierten Aussagen eines psychologischen Tests ersetzen ihrerseits nicht den Diagnostiker.

Testverfahren sind keine 'unabhängige' Instanz, kein 'deus ex machina' : sie sind ebenso 'g e m a c h t', ebenso dem 'bias' ihres Herstellers ausgesetzt, wie das durch 'Erfahrung' mitbestimmte psychologische Gutachten. Tests haben, so gesehen, nur den Vorteil, daß sich das 'bias', durch welches sie mit konstituiert werden, bestimmen läßt und so seinerseits kontrollierbar ist.

Diese Überlegungen dienten der Aufgabe, die mögliche Funktion der Wunschprobe als eines projektiven Verfahrens zu bestimmen.

Die Wunschprobe eignet sich unbestreitbar für jene Zwecke, die wir als explorative oder - weiter gefaßt - als hermeneutische charakterisierten. Aus den Testprotokollen der Versuchspersonen unseres Experiments könnten wir eine Fülle von Beispielen bringen : diese Funktion der Wunschprobe offenkundig zu machen, war indes nicht Aufgabe dieser Arbeit.

Unsere Arbeit stellt vielmehr die Frage, ob die Probleme der Funktionsweise der Wunschprobe und das heißt die Probleme ihrer 'Funktionsebenen' zu umfangreich sind, als daß der - fragliche - Erfolg einer Validierung den Aufwand lohnte, oder ob die differentielle Validierung ein Meßinstrument hervorbringen könnte, welches - seiner Bedeutung für die praktische Diagnostik, der Vielfalt der (dann verläßlich) indizierten Inhalte wegen - derartige 'arbeits-ökonomischen' Erwägungen ausschließen würde. Wir glauben, daß der Aufwand einer differentiellen Validierung allein dann gerechtfertigt ist, wenn in weiteren Untersuchungen geklärt worden wäre, daß das durch diesen Test Indizierte n i c h t für verschiedene Probanden in eine Vielfalt (qualitativ) verschiedener Dimensionen zu zergliedern ist, sondern wenn gezeigt werden könnte, daß die verschiedenen need-scores bei v e r s c h i e d e n e n Versuchspersonen jeweils als Indikator e i n e r bestimmten Dimension (sprich 'Ebene') fungieren.

Die einzige Antwort, die diese Arbeit geben kann, ist die, daß die Wunschprobe - angesichts der aufgewiesenen Vielfalt möglicher Zuordnungen von Index und Indiziertem - in der praktischen Diagnostik nicht verwendet werden kann, a l s o b sie in irgend einer Hinsicht valide wäre. Als hermeneutisches Verfahren jedoch kann sie, so glauben wir, in einzelnen Fällen eine wichtige Rolle spielen. Die Entscheidung, ob sie auch als M e ß instrument für eine Vielzahl verschiedener Variablen brauchbar sein wird, muß weiteren Voruntersuchungen vorbehalten bleiben.

A N H A N G

1. Zur Wunschprobe

a) Das Auswertungsschema

Wir stellen im folgenden für jede einzelne Variable einen repräsentativen Katalog von Testantworten zusammen. Die 'Begründungen' in den Testprotokollen werden mit den einzelnen Variablen signiert, wenn sie sich ihrem sprachlichen Gehalt nach den hier angegebenen Beispielen zuordnen lassen. Bei einigen Variablen war es nötig, die Antworten-Beispiele durch genauere Hinweise zu präzisieren.

Für eine ausführliche Diskussion dieses Auswertungsschemas ist auf Kap. VI, S. 192ff zu verweisen. Formale Gesichtspunkte für die Signierung von Testantworten geben wir im folgenden Abschnitt (s.S.320).

+ n Aggression

aggressiv, frech, gefährlich, wild, bissig, ungebärdig, brutal, stürmisch sein wollen

erschrecken, bedrohen, irreführen, knechten wollen

Raubtier, Jagdvogel, Teufelsmaske sein, Stacheln, Dornen haben wollen

Kritik, Spott, Ironie, Satire repräsentieren wollen

zusehen wollen, wie andere sich abmühen müssen, sich verirren, abstürzen

das Böse ausleben, das Brutale kennenlernen, gegen Gefahren kämpfen wollen

anderen ins Gesicht fliegen wollen (als Seesand)

- n Aggression

nicht zerstören, töten, rauben, erschrecken, kränken, verletzen wollen

nicht gefährlich, feindlich sein wollen

nicht Kummer bereiten, Schaden anrichten wollen

+ n Dominance

Macht, Gewalt, Kraft, Einfluß haben wollen

bestimmend wirken, Ehrfurcht einflößen, hoheitsvoll sein wollen

andere als abhängig erleben, an sich fesseln wollen

- n Dominance

nicht zwingen, gewalttätig, zu machtvoll, zu übermächtig sein wollen

nicht diktatorisch sein, nicht in Ehrfurcht versetzen wollen

n Autonomy

an sich selbst etwas sein, an sich selbst Wert haben, nicht abhängig, nicht an Zweck oder Funktion gebunden sein wollen, nicht von Sympathie abhängig, nicht nur Bestandteil sein wollen

nicht an Schienen, Steuermann, Hersteller (Schriftsteller), Benutzer (Spieler, Leser), Prinzip, Medium gebunden sein wollen

nicht festgelegt sein wollen, nicht sich anpassen müssen wollen

frei sein, wie ein Tier in der Wildbahn, ungebunden sein

Einschränkung :

'nicht Mittel sein wollen' : dieses wird nur dann als 'Aut' signiert, wenn es um den Charakter als M i t t e l geht, wenn es also prinzipiell abgelehnt wird, ein Mittel zu sein;
es wird nicht als 'Aut' signiert, wenn spezifisch eine bestimmte Art von Mittel zu sein abgelehnt wird (z.B. Mittel zur Verschönerung eines anderen), da hier der Akzent der Reaktion sich auf den spezifischen Zweck (hier Verschönerung) richtet

'nicht benutzt werden wollen' : wird darin ausdrücklich ein Abhängigkeitsverhältnis abgelehnt, so wird mit 'Aut' signiert;
kommt jedoch nur die Belanglosigkeit des Gegenstandes zum Ausdruck, wird nicht signiert

+ n Exhibition

schön (nicht häßlich) sein, beachtet, bewundert werden wollen

von Bedeutung sein, wertvoll, wichtig, etwas Außerordentliches sein, Denkmalswert haben wollen

nicht einer unter vielen, nicht zu klein, nichtssagend, zu bescheiden, alltäglich sein wollen, eine persönliche Note haben wollen

für andere interessant sein, gelesen werden, auf andere wirken wollen

nicht mit Widerwillen, Abneigung betrachtet oder benutzt werden wollen

nicht vergessen werden wollen

Einschränkung :

'aus ästhetischem Empfinden heraus' : es wird hier (und in entsprechenden Fällen) nicht mit +Exh signiert : es ist zwar damit nicht eindeutig angezeigt, daß nicht auch eine Identifikation vorliegen könnte, doch scheint in erster Linie mit ästhetischen Wertungen auf den Begriff reagiert worden zu sein, was besagt, daß es sich eher um eine bloße Betrachtung 'von außen' handelt

Problem :

Hier ist es häufig schwierig, zu entscheiden, ob tatsächlich Identifikation vorgelegen hat; dies besonders bei Vpn, die durchgehend knapp, oft nur mit Begriffen, nie aber mit ausdrücklicher Ich-Bezogenheit reagieren. Z.B. :

+'Blumenvase' : 'die Blumen bringen mich zur Geltung' : + Exh

+'Geige' : 'wegen der schönen Töne' : /

- n Exhibition

nicht zu repräsentativ, nicht nur angestarrt werden, nicht auffallen wollen

+ n Nurturance

Anderen helfen, etwas geben wollen, Zuflucht, Geborgenheit, Halt, Trost, Ruhe, Arbeit geben wollen

andere erfreuen, anregen, schützen wollen

anderen dienen, ihnen Schönes, Wissen vermitteln wollen

Einschränkung :

a) Wird nicht deutlich, ob sich die Tendenz der Antwort auf einen anderen Menschen oder auf die Person selbst richtet, wird nicht signiert. Z.B. :

+ 'anderen Ruhe bieten' : + Nur
+ 'Ruhe' : /

b) Es wird unterschieden zwischen 'Freude bringen' und 'Anlaß für Freude sein' :

+ 'da kann ich Menschen erfreuen' : + Nur
+ 'viele freuen sich an mir' : + Exh

Problem :

'bin ein Zufluchtsort' : hier kann es sich um eine Projektion dessen handeln, daß man selbst einen solchen braucht; oder es handelt sich um das Bedürfnis, für andere ein solcher zu sein. Also : entweder 'Suc' oder 'Nur'; wird daher nicht signiert.

- n Nurturance

Ablehnung, Kranke zu transportieren, zu Unglücksfällen unterwegs zu sein, sich aufopfern zu müssen (nur bei 'Sanitätsauto')

+ n Succorance
---

freundliche Zuwendung erfahren, verwöhnt, geliebt werden, gut behandelt werden wollen

+ n Affiliation

Menschen kennen lernen, viel Leben um sich haben, nicht einsam sein wollen

verbindend sein, Zuneigung, Liebe symbolisieren wollen

Geselligkeit, Gemeinschaft, menschliche Wärme erfahren wollen

Einschränkung :

+'Menschen beobachten wollen' : wird nicht mit 'Aff' signiert, da es sich hier nicht ausdrücklich um die Verbindung mit anderen Menschen handelt, sondern eher um beobachtende Distanz, die Erweiterung von Menschenkenntnis

+ n Rejection
---

einsam, uneinnehmbar, undurchlässig, abschirmend, nicht gesellig sein wollen

- n Rejection
---

gegen Entfremdung, Isolation, Distanz, Abwehr, das Unpersönliche unter den Menschen sein

+ n Deference
---

seinen Zweck, eine Aufgabe erfüllen, von daher einen Sinn haben, sich einem größeren Zweck unterordnen wollen

- n Deference
---

nicht jemandem dienen, nicht untergeordnet sein wollen, nicht Rahmen, Kulisse, Attribut sein wollen

n harmavoidance
---

nicht zerstört, verzehrt, abgebrochen, zertreten, schlecht behandelt werden, nicht vergänglich sein wollen

allgemein ausgedrückte Furcht vor nicht näher bestimmten Dingen, vor Gefahr, Leiden, Krankheit, Unheimlichem, Bedrückendem

b) Formale Hinweise für die Signierung

Jede Test-Antwort (Begründung einer item-Wahl) wird daraufhin analysiert, ob sie sich dem sprachlich-manifesten Gehalt nach einer der need-Kategorien zuordnen läßt.

Test-Antworten, bei denen eine klare Kategorisierung oder überhaupt eine Zuordnung zu 'needs' nicht möglich ist, werden nicht signiert.

Beispiel :

(+) Geige : "wegen der schönen Töne" : /
(+) Lehrbuch: "weil es doch viel Information enthält" : /[1)]

Die zwei Antworten, die gegeben werden, wenn ein item in die +/- Kategorie gelegt wird, werden b e i d e mit einem ganzen score signiert : es liegt kein Grund dafür vor, den Teilen einer Doppelantwort psychologisch ein geringeres Gewicht zuzuschreiben als einer einzelnen Antwort auf ein item.

Beispiel :

(+/-) Nordsee : + "das Mächtige, Naturgewalten :+nDom
- "das für den Menschen Gefährliche, ertrinken" :-nAgg
(+/-) Orkan : + "daß ich mich austoben könnte" :+nAgg
- "daß ich zu viel Unheil anrichte" :-nAgg

Läßt sich eine längere Antwort zwei need-Kategorien zuordnen, werden beide signiert :

Beispiel :

(+) Dornenhecke : "man kann die Leute stechen, wenn man sie nicht will" :+nAgg :+nRej
(+) Orkan : "ich würde die Menschen erschrecken, und brauchte mich nicht zu beugen" :+nAgg :+nAut

---

1) Vor dem item wird jeweils angezeigt, welcher Antworten-Kategorie (+, -, +/-) die angeführte Begründung zuzuordnen ist.

Stellt eine längere Antwort die mehrfache Variation desselben Gesichtspunktes dar, wird nur einmal signiert :

Beispiel :

(-) Orkan : :"sehr gefährlich, bedroht die Menschen, müssen sich vor ihm in acht nehmen" :-nAgg

Tauchen in einer Antwort Gesichtspunkte auf, die nicht direkt zur Begründung der item-Wahl dienen, etwa in einer +Antwort einzelne negative Gesichtspunkte, so werden sie dann signiert, wenn sie als Ausnahme von der positiven Begründung ausgenommen werden bzw. wenn sich die Antwort in eine $^{+}/_{-}$Antwort umformen ließe; sie werden nicht signiert, wenn sie ausdrücklich als irrelevant bezeichnet werden, also keine Abschwächung der Wahl bedeuten.

Beispiel :

(-) Grashalm : "nur auf einer schönen Wiese, wo nicht alle herumtreten; Freiheit, Ungezwungensein" : avoid/+nAut

Kommt eine Variable nicht direkt zum Ausdruck, aber wird die Antwort durch das zugehörige item vereindeutigt, so wird entsprechend signiert :

Beispiel :

(+) Orkan : "hätte sehr viel Kraft, hab' dann ganz schöne Spuren hinterlassen" : +nAgg

Teilweise läßt sich auch die Richtung der Antwort nur durch das zugehörige item entscheiden :

Beispiel :

(-) Tiger : "ist zu gefährlich"
man möchte für andere nicht gefährlich, also nicht aggressiv sein :-nAgg

(-) Auto : "ist zu gefährlich"
man käme selbst in gefährliche Situationen, die man vermeiden möchte :avoid.

c) Tabelle : die Auswertung der Wunschprobe am Beispiel der need-scores einzelner Vpn (in %-scores)

Variable	G_1	Ph_2	Sz_1	Ph_{11}	Ps_6	Ps_1	Ps_9
-Agg	13,5	3,5	12,5	2,3	25,7	30,0	9,1
+Agg	8,1		2,5	22,8			
-Agg/+Agg	1,7		5,0	0,1			
+Dom	18,9	3,5		9,3	8,6		
-Dom	2,7						
+Def	8,1		2,5				
-Def	5,4					4,4	3,0
+Nur	16,2	3,5	30,0	4,6	25,7	13,0	18,2
+Suc							
+Aff	2,7	6,9			8,6	4,4	24,2
+Rej		3,5	2,5				
-Rej	2,7				5,8		3,0
+Aut	5,4	34,5	12,5	23,1	2,9	8,8	42,5
+Exh	13,5	13,8	17,5	18,6	17,1	30,0	
avoid	2,7	27,6	20,0	13,8	5,8	8,8	

2. Zum Q_s-sort-Verfahren

a) Gesichtspunkte für die item-Konstruktion

Die Variablen, welche das Q_s-sort-Verfahren enthalten soll, sind vorgegeben : sie wurden nach den Erfordernissen der WP-Analyse ausgewählt (vgl. S.127).

Von diesem Ausgangsmaterial abgesehen, müssen die verwendeten Variablen inhaltlich hinreichend heterogen sein : denn die items sind einer Skala zuzuordnen, welche ein Kontinuum von 'hohem Zutreffen' bis zu 'starker Ablehnung' beschreibt. Das Q_s-sort muß also so zusammengestellt sein, daß jede Vp positive wie negative Wahlen treffen kann. Die der WP entnommenen Variablen werden daher für das Q_s-sort-Verfahren durch 'komplementäre' Variablen ergänzt (vgl. S.146).

Jede Variable ist durch hinreichend viele items zu repräsentieren : denn jede Variable soll für sich genommen, d.h. abgelöst von den übrigen Variablen, betrachtet werden können. Für jede need-Kategorie sind daher verschiedene Einschätzungen zu erheben, die sich in ihrem Mittelwert ('zentrale Tendenz') zusammenfassen lassen (vgl. S.204).

Die items einer Variablen sind - nach Möglichkeit - vielfältig genug zu konzipieren : denn derselbe need kann sich, intra-individuell gesehen, in verschiedenen Situationen verschieden stark manifestieren; entsprechend bestehen inter-individuelle Differenzen darin, in welchen Situationen ein need zum Ausdruck kommt (vgl. S.147).

Zugleich ist ein einzelnes item allgemein genug zu formulieren : denn jede Vp soll das angegebene Verhalten auf sich beziehen können. Es sind also allgemeine Situationen wie 'Diskussion', 'Studienfach', 'Freunde', 'Pflichten' etc. zur Beschreibung des Verhaltens zu verwenden.

Die items sind - nach Möglichkeit - auf der Ebene des V e r h a l t e n s zu formulieren. Denn die Vp soll einmal ihr 'tatsächliches Verhalten' einschätzen; zum anderen soll sie

anhand derselben items angeben, ob und wieweit das in den items beschriebene Verhalten ihren Ich-Ideal-Normen entspricht; und sie soll anhand derselben items darstellen, wie sie sich v e r h a l t e n würde (d.h. in welchem Maße sie durch die items charakterisiert wäre), wenn sie ihre Wünsche verwirklichen könnte und nicht sozialen Normen zu entsprechen hätte.

Jedes item muß so formuliert werden, daß es einem Kontinuum verschiedener Grade des 'Zutreffens', bzw. einem Kontinuum verschiedener 'Ausprägungsgrade' des beschriebenen Verhaltens zugeordnet werden kann. Das bedeutet : bei jedem item muß die Vp entscheiden können, ob es der I n t e n s i t ä t oder der H ä u f i g k e i t nach variiert. Die items müssen - allgemein gesagt - adverbiale Bestimmungen enthalten, die sich in eine Abstufung von 'mehr' oder 'weniger' übersetzen lassen [1].

Die items sind so zu konzipieren, daß sie sich unabhängig voneinander auffassen und einschätzen lassen. Es sind allerdings bisher keine befriedigenden Kriterien entwickelt worden, wie diese Forderung praktisch erfüllt werden kann. Für eine ausführliche Diskussion sei hier auf den EXKURS verwiesen (vgl. S. 171f).

Als S t i l der Formulierung empfiehlt sich eine (konstatierende) Verhaltensbeschreibung. Es ließ sich zeigen, daß andere Stilformen entweder defensiveness auslösen, oder dazu

---

1) Das item 'Es fällt mir leicht, andere zu überzeugen' variiert z.B. nach der 'Intensität', da es dem Kontinuum von 'Es fällt mir sehr leicht' bis 'Es fällt mir sehr schwer' zugeordnet werden kann. Das item 'Ich bin immer bereit, anderen etwas auszuleihen' variiert nach der 'Häufigkeit' : es läßt sich dem Kontinuum von 'immer' bis 'selten' oder 'nie' zuordnen.

führen, daß die entsprechenden items bevorzugt als 'zutreffend' eingeschätzt werden [1].

b) Die items des Q_s-sort-Verfahrens

Wir geben im folgenden eine Zusammenstellung der 88 items, welche, auf einzelnen Karten (3 x 9 cm) abgedruckt, durch die Vpn in jeweils zufälliger Reihenfolge 'sortiert' werden. Den 8 items einer Variablen stellen wir einen Ausschnitt aus der Murray'schen Definition der betreffenden Variablen voran und zwar jenen Teil der sehr umfangreichen phänomenalen Beschreibung einer Variablen, den Murray als 'desires and effects' überschreibt. - Zur genaueren Bestimmung einer Variablen übernehmen wir von Murray noch einen weiteren Gesichtspunkt : wir vermerken auch jene need-Kategorien, mit welchen die besprochene Variable u.U. gemeinsam auftritt (fusion). 'Verhalten' läßt sich nicht immer eindeutig einer need-Kategorie zuordnen : so kann etwa Aggression u.a. verschmelzen mit nDom, nAuto, nExh. Diese Angaben sind wichtig, da es sich auch bei der Konstruktion der items nicht vermeiden läßt, daß das (durch das item) beschriebene Verhalten zugleich als Manifestation einer anderen Variablen aufgefaßt werden kann. Unter der Bezeichnung 'fusions with' gehen wir allerdings nur die Variablen an, die in unserer Untersuchung verwendet werden.

---

1) Buss (1959) verglich 7 verschiedene item styles, z.B. Satzeinleitungen wie 'I must admit ...' oder 'I can't help ...' oder 'Like most people ...' und stellte fest, daß ein derartiger sprachlicher Kontext des beschriebenen Verhaltens sich auf die Art der Selbsteinschätzung der betreffenden items nachweislich auswirkt.

n A g g r e s s i o n (I)

Desires and Effects : Physical : To overcome opposition forcefully. To fight. To revenge an injury. To attack, injure or kill an O. To oppose forcefully or punish an O.[1] Verbal : To belittle, censure, curse or ridicule maliciously an O. To depreciate and slander (a.a.O. S.159).

Fusions with : nDom, nAuto, nExh

1. Selbst Kleinigkeiten können mich manchmal so erregen, daß mein Verhalten gar nicht mehr dem Anlaß angemessen ist.
2. Ich gerate leicht in Ärger und halte ihn dann auch nicht zurück.
3. Vorwürfe, welcherart auch immer, weise ich stets zurück.
4. Wenn Dinge nicht so klappen, wie geplant, lasse ich leicht den Ärger darüber an irgend einer Person oder Sache aus.
5. Es liegt mir mehr, bei einem Streit Partei zu ergreifen, als ausgleichend zu wirken.
6. In Diskussionen werde ich leicht scharf, um mich zu behaupten.
7. Es macht mir nichts aus, andere zu verletzen, wenn es darum geht, Klarheit zu schaffen.
8. Anschuldigungen, auch wenn sie zutreffen, kann ich nicht ertragen : ich werde eher gereizt und heftig.

n A b a s e m e n t (II)

Desires and Effects : To submit passively to external force. To accept injury, blame, criticism, punishment. To surrender. To become resigned to fate. To admit inferiority, error, wrong-doing or defeat. To confess and atone. To blame, belittle or mutilate the self. To seek and enjoy pain, punishment, illness and misfortune. (a.a.O. S.161f)

Fusions with : nSuc, nExh, nDef

---

1) Das Zeichen 'O' steht hier und im folgenden für 'object', i.e. : any external entity (thing, animal, person) other than the subject.

1. Bei einem Streit bestehe ich nicht lange auf meinem Recht, sondern gebe es eher auf.
2. Dinge, die sich mir in den Weg stellen, empfinde ich meist als unabwendbar und versuche nicht erst, sie aus dem Wege zu räumen.
3. Ich bringe es kaum fertig, andere zu blamieren und schreibe dann lieber mir die Schuld zu.
4. Wenn Leute mich kritisieren, lasse ich sie gewähren und leiste kaum Widerstand.
5. Die Anwesenheit von Leuten, die mir überlegen sind, macht mich ziemlich unsicher.
6. Ich habe oft das Gefühl, daß ich ungerecht behandelt werde.
7. Ich fühle mich den meisten Altersgenossen unterlegen.
8. Häufig befällt mich ein Gefühl des Überflüssigseins.

n Dominance (III)

Desires and Effects : To control one's human environment. To influence or direct the behavior of Os by suggestion, seduction, persuasion, or command. To dissuade, restrain, or prohibit. To induce an O to act in a way which accords with one's sentiments and needs. To get Os to co-operate. To convince an O of the 'rightness' of one's opinion. (a.a.O. S.152)

Fusions with : nAgg

1. Ich ertrage nur schwer die Erfahrung, hinter anderen zurückzustehen.
2. Ich habe ein starkes Bedürfnis nach einem Betätigungsfeld, in dem ich Einfluß gewinnen kann.
3. Häufig bin ich derjenige, der andere zu einer Entscheidung führt.
4. Wenn ich einen neuen Menschen kennenlerne, versuche ich meist von mir aus das Gespräch und die Situation zu lenken und zu prägen.
5. Wenn es etwas zu planen, zu organisieren oder durchzusetzen gilt, bin ich sogleich daran engagiert.
6. Selbstbestätigung finde ich vor allem darin, daß ich durch meine Person auf andere wirke.
7. Es liegt mir sehr, mich innerhalb einer Gruppe zu deren Sprecher zu machen und neue Aktivitäten und Initiativen zu planen.
8. In Diskussionen bin ich nur zögernd bereit, mich den Argumenten anderer unterzuordnen.

n D e f e r e n c e (IV)

Desires and Effects : To admire and support a superior O. To praise, honour, or eulogize. To yield eagerly to the influence of an allied O. To emulate an exemplar. To conform to custom. (a.a.O. S. 154)

Fusions with : nSuc, nAba

1. Der Ratschlag älterer Leute ist mir sehr wichtig.
2. Ich halte mich lieber im Hintergrund auf, als daß ich gern in den Mittelpunkt geriete.
3. Ich gewinne Selbstsicherheit dadurch, daß ich mich überall leicht unterordnen kann.
4. Es macht mich sicher, wenn ich mich an allgemeine Gepflogenheiten halten kann.
5. Oft bemerke ich, daß ich jemanden nachahme oder mit jemandem in meinen Ansichten übereinstimme, den ich als mir überlegen ansehe.
6. Ich orientiere mich gern am Vorbild nachahmenswerter Leute.
7. Von Leuten, die über mir stehen, sehe ich eher ihre guten, als ihre schlechten Seiten.
8. Meine Aufgaben erfülle ich selbstverständlich so, wie man es von mir erwartet.

n N u r t u r a n c e (V)

Desires and Effects : To give sympathy and gratify the needs of a helpless O : an infant or any O that is weak, disabled, tired, inexperienced, infirm, defeated, humiliated, lonely, dejected, sick, mentally confused. To assist an O in danger. To feed, help, support, console, protect, comfort, nurse, heal. (a.a.O. S.184)

Fusions with : nAff, nDom, nDef, nAba, nAgg

1. Ich verwende gern Zeit und Energie darauf, wenn andere mich um etwas bitten.
2. Wenn ich andere in Bedrängnis sehe, werden mir meine eigenen Belange ziemlich unwichtig.
3. Ich nehme gern einigen Ärger auf mich, wenn ich dafür anderen behilflich sein kann.
4. Zu Menschen, die weniger glücklich sind als ich, fühle ich mich hingezogen.
5. Wenn ich etwas geschenkt bekomme, möchte ich immer andere daran teilhaben lassen.

6. Menschen, die traurig und niedergeschlagen sind, kann ich nicht aus dem Wege gehen, sondern ich muß versuchen, sie zu ermutigen.
7. Es fällt mir schwer, über andere etwas Schlechtes zu sagen.
8. Ich bin immer bereit, anderen etwas auszuleihen.

n Succorance (VI)

Desires and Effects : To have one's needs gratified by the sympathetic aid of an allied O. To be nursed, supported, sustained, surrounded, protected, loved, advised, guided, indulged, forgiven, consoled. To remain close to a devoted protector. To have always a supporter. (a.a.O. S. 182)

Fusions with : nHarm, nAff, nExh, nAba, nDom

1. Gewöhnlich brauche ich jemanden, mit dem ich über persönliche Schwierigkeiten oder ein Mißgeschick sprechen kann.
2. Unerwartete Schwierigkeiten entmutigen mich schnell.
3. Ich brauche es, daß andere mir ihre Sympathie bekunden.
4. Das Mitgefühl anderer ist mir wie eine gute Medizin.
5. Wenn ich aus eigener Verantwortung handeln soll, werde ich sehr unsicher.
6. Ich bin abhängig davon, daß sich jemand um mein Wohlergehen sorgt.
7. Es bedeutet mir sehr viel, wenn andere mir ihre Unterstützung zusagen.
8. Es gelingt mir schwer, mich selber wieder aufzurichten : ich brauche den Trost anderer.

n Affiliation (VII)

Desires and Effects : To draw near and enjoyably co-operate or reciprocate with an allied O : an O who resembles the S or likes the S (the subject, Verf.). To please and winn affection of a cathected O. To adhere and remain loyal to a friend. (a.a.O. S. 174)

Fusions with : nNur

1. Ich habe häufig den Wunsch, mit anderen etwas Gemeinsames zu unternehmen.
2. Es fällt mir nicht schwer, andere Dinge liegen zu lassen, damit ich mit meinen Freunden zusammen sein kann.
3. Ich arbeite lieber in einem Team als für mich alleine.

4. Ich finde überall sehr schnell Freunde.
5. Oft tue ich Dinge nur deshalb, weil andere sich daran freuen.
6. Einladungen zu Bekannten kann ich meist kaum widerstehen.
7. Ich vermeide es meist, andere zu kritisieren und ihnen entgegenzustehen.
8. Ich bin sehr unglücklich, wenn ich nicht unter Menschen bin.

n R e j e c t i o n (VIII)

Desires and Effects : To separate oneself from a negatively cathected O. To exclude, abandon, expel, or remain indifferent to an inferior O. To snub or jilt an O. (a.a.O. S. 177)

Fusions with : nAuto, nAgg

1. Es macht mir nicht viel aus, Freundschaften abzubrechen.
2. Ich werde leicht ungehalten, wenn ich mit Leuten zusammentreffe, deren Meinungen schwach und unbegründet sind.
3. Oft vermeide ich es, mit irgendwelchen Bekannten zusammenzutreffen.
4. Fremden gegenüber bin ich zunächst sehr reserviert.
5. Für Leute, die mich langweilen, ist mir meine Zeit meist zu schade.
6. Ich bin wenig tolerant, wenn andere nicht meinen Geschmack teilen.
7. Ich bin besonders kritisch in der Auswahl meiner Freunde.
8. Am liebsten lebe ich abgeschlossen und zurückgezogen von anderen.

n A u t o n o m y (IX)

Desires and Effects : To get free, shake off restraint, break out of confinement. To resist coercion and restriction. To avoid or quit activities prescribed by domineering authorities. To be independent and free to act according to impulse. To be unattached, unconditioned, irresponsible. To defy conventions. (a.a.O. S. 156)

Fusions with : nAgg, mRej, nExh, nDom

1. Ich traue mir zu, daß ich in meinem Fach einen unabhängigen Standort erarbeite.
2. Ich leiste deutlichen Widerstand, wenn andere mich in meinem Tun irgendwie einschränken wollen.
3. Ich lasse mich nicht gern zu irgendwelchen Dingen verpflichten.
4. Die Meinungen anderer interessieren mich wenig, lieber gehe ich meinen eigenen Weg.
5. Ich habe einen ausgeprägten Willen, meine Unabhängigkeit deutlich zu machen und durchzusetzen.
6. Autoritätspersonen reizen mich immer dazu, sie zu kritisieren.
7. Welche Regeln und Gepflogenheiten für mich verbindlich sind, lasse ich mir durch niemanden vorschreiben.
8. Es ist mir wichtig, den Anschein zu vermeiden, daß ich mich irgendwelchen Konventionen anpasse.

n E x h i b i t i o n (X)

Desires and Effects : To make an impression. To be seen and heard. To excite, amaze, fascinate, entertain, shock, intrigue, amuse or entice Os. (a.a.O. S. 170)

Fusions with : nAff, nDom, nSuc

1. Ich stelle mir öfters vor, wie reizvoll es wäre, in irgendeiner Weise Aufsehen zu erregen.
2. In meinen Wunschträumen identifiziere ich mich gern mit Menschen, die in der Öffentlichkeit angesehen sind.
3. Es befällt mich ein Gefühl der Unzufriedenheit, wenn man nicht Notiz von mir genommen hat.
4. Es gibt mir eine außerordentliche Befriedigung, wenn ich spüre, wie andere mich bewundern.
5. Ich versuche immer wieder, die Augen anderer auf mich zu lenken.
6. Ich brauche es, daß meine Leistungen von anderen gebührend wahrgenommen werden.
7. Ich nehme gerne die Gelegenheit wahr, mich selbst auf irgendeine Weise darzustellen.
8. Ich würde gern mit dem Gefühl leben, unersetzbar zu sein.

n H a r m a v o i d a n c e (XI)

Desires and Effects : To avoid pain, physical injury, illness and death. To escape from a dangerous situation. To take precautionary measures. (a.a.O. S. 197)

Fusions with : -

1. Meine Gedanken beschäftigen sich oft damit, von wieviel Gefahren man umgeben ist.
2. Der Gedanke der eigenen Vergänglichkeit belastet mich sehr.
3. Oft verfolgt mich der Gedanke, es könnte mich irgendein Schicksalsschlag treffen.
4. Ich scheue mich ziemlich davor, allein in einer fremden Stadt zu sein.
5. Ich habe manchmal das Gefühl, daß mich jemand von hinten angreifen könnte.
6. In Situationen, die ich nicht ganz überschauen kann, habe ich leicht ein Gefühl der Angst.
7. Es befällt mich häufig die Sorge, daß ich etwas Schreckliches erleben müßte.
8. Vor Fahrten muß ich immer daran denken, daß ein Unfall passieren könnte.

3. Instruktionen

a) Allgemeine Vorinstruktion

Es geht in dieser Untersuchung um die Analyse verschiedener psychologischer Verfahren, deren Funktionsweise noch nicht hinlänglich bekannt ist. Es soll festgestellt werden, wo die Schwierigkeiten dieser Methoden liegen, wo sie etwa besser ausgearbeitet werden müssen, oder was an ihnen besonders zu berücksichtigen ist.

Ihre Mitarbeit dient also einer solchen rein wissenschaftlichen Fragestellung. Es mag Sie daher vielleicht verwundern, wenn die Verfahren, die ich Ihnen vorlegen werde, Sie doch ganz persönlich ansprechen sollen. Aber das liegt eben in der Natur d e r Verfahren, die hier zur Untersuchung stehen.

Ihr Name wird auf jeden Fall auf dem Testmaterial verschlüsselt, so daß es weder mir noch anderen möglich sein wird, die Antworten später einer bestimmten Person zuzuordnen; es sei denn, Sie hegten selber den Wunsch, noch einmal Einblick in Ihr eigenes Material zu nehmen.

Aus diesem Grunde können und sollen Sie sich völlig frei fühlen, alle Fragen spontan und offen zu beantworten : je mehr Bereitschaft Sie mitbringen, in dieser Form ernsthaft mitzuarbeiten, desto mehr dient das der Sache; je freier Sie auf ein Verfahren reagieren, je mehr Erkenntnisse können über das Verfahren gewonnen werden; je mehr Sie dabei vor sich selbst ehrlich sind, desto eher ist die Untersuchung auch für Sie persönlich interessant. Überlassen Sie sich also ganz frei dem Material, das ich Ihnen vorlegen werde.

Sie werden es im Verlauf der 2 Sitzungen mit verschiedenen derartigen Verfahren zu tun haben, und ich werde Ihnen dabei im einzelnen Fall genau schildern, was Ihre Aufgabe sein wird. Ich möchte Sie schon jetzt bitten, diese Erklärungen genau zu verfolgen und anschließend ebenso genau zu beherzigen, damit in dieser Untersuchung wirklich ein bestmöglicher Effekt erzielt werden kann.

Meine letzte Bitte ist nun, daß Sie mir im Anschluß an jedes Verfahren ein wenig Ihre Eindrücke schildern, über etwaige Schwierigkeiten oder Auffälligkeiten berichten. Über die Untersuchung insgesamt werden wir dann erst nach Abschluß der zweiten Sitzung sprechen. Ich bin dann auch gerne bereit, Ihnen eingehender Auskunft zu geben über diesen Bereich der Psychologie.

b) Instruktion zur Wunschprobe

Die erste Aufgabe wendet sich an Ihre Fantasie. Stellen Sie sich vor, Sie würden in die Lage versetzt, nicht als Mensch zu leben, sondern wählen zu können, was Sie stattdessen sein möchten.

Diese Ihre Wahl ist nun begrenzt durch diese Kärtchen, die ich Ihnen hier vorlege, z.B. 'Baum' : möchten Sie ein Baum sein? Sie haben nun 3 bzw. 4 Möglichkeiten der Wahl. Sie möchten etwas sehr gerne sein (dann legen Sie die Karte bitte an diese Stelle). Sie möchten etwas eindeutig nicht sein (dann ....). Eine Karte wirkt auf Sie sowohl positiv wie im gleichen Maße negativ, so daß Sie sich nicht zwischen beiden entscheiden können (dann ....). Es kann auch einmal vorkommen, daß eine Karte Sie überhaupt nicht anspricht, so daß Sie weder ja noch nein zu dem bezeichneten Gegenstand sagen können (sie wäre dann hier abzulegen). Von dieser Form sollen Sie allerdings möglichst geringen oder gar keinen Gebrauch machen.

Wichtig ist nun für Sie zu wissen, daß es auf keinen Fall eine irgendwie bevorzugte Antworten-Kategorie gibt. Sie betrachten jede Karte einzeln und legen so viele Karten in eine Antworten-Kategorie, wie es Ihrem Wunsch entspricht. Zu beachten ist weiter, daß es unter diesen Karten nicht etwa irgendwelche logischen Verbindungen gibt. Sie beantworten jede Karte für sich und spontan. Denken Sie nicht bei einer Karte, Sie hätten eine ähnliche schon einmal beantwortet und müßten es nun in derselben Weise tun. Auch ähnliche Karten können ganz verschiedene Vorstellungen hervorrufen, und ganz verschiedenartige Wünsche können nebeneinander bestehen. 'Psychologisches' ist nicht unbedingt 'logisch'.

Wählen wir einmal eine Karte - 'Baum' : möchten Sie ein Baum sein? - (Vp antwortet.) - Würden Sie nun bitte sagen, aus welchem Grunde Sie ein 'Baum' sein möchten? Antworten Sie dabei bitte ganz spontan, ohne erst lange darüber nachzudenken. Versuchen Sie ganz unmittelbar zum Ausdruck zu bringen, was Sie bei Ihrer Entscheidung bewogen hat, was den Ausschlag für Ihre Wahl gab.

Nehmen Sie nun bitte eine Karte nach der anderen und lesen Sie sie bitte jeweils laut vor. Wenn Sie Ihre Wahl getroffen haben, sagen Sie mir bitte, wie in diesem Beispiel, was Sie zu dieser Wahl bewog. Denken Sie bitte daran : Sie sollen überlegen, ob Sie das, was die Karte beschreibt, s e i n möchten.

c) Instruktion zum Q_s-sort-Verfahren

In unserer nächsten Aufgabe haben Sie es wieder mit einzelnen Karten zu tun, die Sie nun jedoch in anderer Weise ordnen sollen. Sie sollen einmal darüber nachdenken, wie Sie sich selbst sehen und erleben. Auf jeder Karte finden Sie einen Satz, einen persönlichkeits-beschreibenden Satz. Sie sollen dann entscheiden, ob dieser Satz Sie so beschreibt, wie Sie sich selbst in Ihrem Verhalten erfahren und einschätzen oder ob dieser Satz gar nicht auf Sie zutrifft, d.h. kaum oder gar nicht für Sie charakteristisch ist.

Es gibt hier nun allerdings nicht nur die Antworten 'ja' und 'nein', sondern Sie sollen in 9 Gradabstufungen kennzeichnen, wie sehr Sie sich durch jeden einzelnen Satz charakterisiert fühlen.

Diese Aufgabe möchte ich Ihnen an einem Beispiel veranschaulichen : z.B. der Satz "Ich gehe oft ins Kino". Wenn Sie besonders häufig ins Kino gehen, sind Sie auch durch diesen Satz besonders zutreffend gekennzeichnet. So werden Sie ihm vielleicht den Punktwert +4 geben. Vielleicht gibt es aber noch eine Reihe anderer Sätze, durch welche Sie sich stärker charakterisiert fühlen, so daß Sie dieser Beschreibung (Sie gingen oft ins Kino) nur den Platz +3 oder auch nur +2 geben.- Sollten Sie nun allerdings überhaupt nicht ins Kino gehen, so sind Sie entsprechend durch diesen Satz überhaupt nicht gekennzeichnet und Sie werden die Karte in die Rubrik -4 legen. Hier kann es nun auch sein, daß es wiederum einige weitere Sätze gibt, welche noch weniger auf Sie zutreffen, so daß diese Karte über den Kino-Besuch vielleicht eher unter den Punkt -1 gehört.

Sie sehen, zum Schluß sollen die Karten so geordnet sein, daß sie, je weniger sie auf Sie zutreffen, desto mehr in dieser Richtung liegen (nach -4 hin), und daß sie, je mehr sie auf Sie zutreffen, in dieser Richtung liegen (nach +4 hin). Sie sehen, es ist dies keine leichte Aufgabe, und Sie müssen bei jeder Karte sehr genau überlegen, in welche Rubrik sie einzuordnen ist.

Es kann sogar sein, daß Sie sich manche dieser Fragen selbst noch nicht gestellt haben, so daß Sie erst einmal darüber nachdenken müssen, wie weit Sie überhaupt durch einen bestimmten Satz charakterisiert sind.

Hier gibt es nun eine Art Hilfestellung, die Ihnen die Entscheidung darüber erleichtern soll, in welchem Grade ein bestimmter Satz auf Sie zutrifft. Sie sollen sich nämlich dabei jeweils überlegen, ob Sie durch einen solchen Satz stärker charakterisiert oder weniger charakterisiert sind als andere Menschen. Wir wählen eine bestimmte Bezugsgruppe, mit der Sie sich vergleichen sollen. Diese Bezugsgruppe sind für Sie Studentinnen der ersten Semester. Sie stellen also fest, in welchem Maße die auf einer Karte wiedergegebene 'Behauptung' für Sie - im Vergleich mit den Gliedern Ihrer Bezugsgruppe - zutreffend ist oder nicht. Das sei ebenfalls an einem Beispiel erläutert.

Wir nehmen noch einmal das Beispiel mit dem Kino. Was heißt : "Ich gehe oft ins Kino"? Für Sie persönlich und innerhalb Ihrer Familie kann es als 'oft' angesehen werden, wenn Sie in jedem Monat einmal ins Kino gehen, statt etwa nur alle halbe Jahr. Im Vergleich zu Ihrer Bezugsgruppe kann es aber umgekehrt als relativ 'selten' gelten, wenn Sie nur einmal im Monat, statt wie die anderen der Gruppe, jede Woche ins Kino gehen. Da Sie sich nun an dieser G r u p p e ausrichten sollen, ist zu folgern, daß Sie durch den Satz "Ich gehe oft ins Kino" nicht besonders charakterisiert sind, da Sie ja - in diesem Beispiel - sehr viel seltener gehen als alle anderen. Sie würden also eine Rubrik auf der negativen Seite zu wählen haben. - Wenn Sie nun in einem anderen Falle nicht mehr und nicht weniger ins Kino gehen, als es für den Durchschnitt Ihrer Gruppe gilt, so wäre das durch eine Zahl im mittleren Bereich dieser Skala darzustellen.

Ihre erste Frage lautet also immer : "Wie sehr trifft dieser Satz auf mich zu". Um dieses entscheiden zu können, fassen Sie gleichzeitig ins Auge, in welchem Maße derselbe Satz die anderen Ihrer Bezugsgruppe charakterisiert. An dem Beispiel, das ich Ihnen eben gab, wurde deutlich, daß die Antwort sehr verschieden ausfallen kann, je nachdem auf welche Gruppe man sich bezieht. Ich erinnere noch einmal daran, daß Sie sich im Vergleich zu gleichaltrigen Studentinnen sehen sollen.

Auch hier ist darauf aufmerksam zu machen, daß es keinesfalls um irgendwelche logischen Zuordnungen geht. Jede Karte ist für sich zu betrachten, und sie ist, je nachdem, ob die Entscheidung sofort eindeutig getroffen werden kann, gegen die anderen abzuwägen und dann einer Rubrik auf der vorliegenden Skala zuzuordnen. Dabei können - u.U. - sehr widersprüchliche Inhalte den gleichen Punktwert erhalten.

Einen wichtigen Hinweis muß ich Ihnen noch geben. Wenn Sie die Karten auf die verschiedenen Rubriken der Skala verteilen, so sind Sie dabei in einer bestimmten Weise festgelegt. Es ist nämlich vorgeschrieben, wieviel Karten zum Schluß in jeder Rubrik liegen dürfen. Ich weiß, daß diese Vorschrift nicht leicht einzuhalten ist. Es ist für Sie erschwerend und auch störend, daß Sie nicht vollkommen frei entscheiden können und daß Sie teilweise die Karten umsortieren müssen, damit die vorgeschriebene Verteilung entsteht. Es ist dies eine Begrenzung, die in der Methode selbst liegt und die nicht aufgehoben werden kann. Anschließend will ich Ihnen gern die Gründe für diese Vorschrift erläutern. Und ich bin interessiert daran, von Ihnen zu hören, welcherart Schwierigkeiten Sie im Umgang mit diesem Verfahren hatten. Die vorgeschriebene Anzahl für jede Rubrik ist oben auf der Skala deutlich vermerkt.

Abschließend möchte ich Sie noch einmal bitten : arbeiten Sie so sorgfältig wie möglich, überprüfen Sie genau, welchen Punktwert Sie jeder Karte zuweisen. - Ihre Aufgabe besteht darin, anhand der vorgelegten Karten zu beschreiben, wie Sie Ihr V e r h a l t e n einschätzen.

Es kann sein, daß Sie teilweise das Gefühl haben, daß Sie sich durch andere Inhalte sehr viel besser charakterisieren könnten; es kann sein, daß Sie bestimmte charakteristische Beschreibungen vermissen. Doch geht es nur darum, daß Sie sich anhand des vorliegenden Materials beurteilen. Das hat zur Folge, daß die Rubriken +4 und -4 nicht beinhalten, was Sie 'überhaupt' am meisten oder wenigsten kennzeichnet. Sie sortieren dort jene Karten ein, welche von den gegebenen die zutreffendsten sind.

Nachdem Sie sich nun gut mit dieser Methode eingearbeitet haben, wollen wir das Ganze noch einmal, aber jetzt in einem anderen Sinne betreiben. Stellen Sie sich jetzt nicht darauf ein, wie Sie sich selbst in Ihrem Verhalten sehen, sondern darauf, was Sie für Ihre Persönlichkeit a n s t r e b e n , was Sie sich für Ihr Verhalten zum Vorbild gemacht haben, also darauf, was Sie von sich selbst erwarten; unabhängig davon, ob Sie es schon erfüllt haben oder nicht. Versuchen Sie also nun nach den gleichen Regeln wie eben Ihr Ich-Ideal darzustellen.

Sie sollen sich bei diesem dritten und letzten Durchgang wieder auf etwas anderes einstellen. Sie sollen nun darstellen, wie Sie sich am liebsten verhalten w ü r d e n , wenn es die Anforderungen der Umwelt, die Normen, die Erwartungen der anderen, bestimmte Pflichten gar nicht gäbe. Versetzen Sie sich einmal in diesen - durchaus irrealen - Zustand, in welchem man allein nach seinen eigenen Wünschen und Bedürfnissen leben könnte. Versuchen Sie sich ganz auf diese Möglichkeit einzustellen und ordnen Sie die Karten danach, wie Sie sich am liebsten verhalten würden, wenn es nur möglich wäre.

d) Instruktion zum Erinnerungsexperiment

Sie werden sich erinnern, daß wir es bei der ersten Sitzung mit 2 verschiedenen Verfahren zu tun hatten, in beiden Fällen hatten Sie mit Karten umzugehen, einmal mit Wörtern, einmal mit Sätzen. Es interessiert mich nun, wieviele der Wörter des ersten Verfahrens Sie behalten haben. (Ihre Aufgabe war es ja gewesen zu entscheiden, ob Sie sich in den durch die Wörter dargestellten Gegenstand verwandeln lassen wollten oder nicht.) Wenn ich Ihnen das Zeichen gebe 'bitte anfangen', dann schreiben Sie bitte der Reihe nach so viele Wörter auf, wie Ihnen nur einfallen.

e) Instruktion zu real-self (2) des Q_s-sort-Verfahrens

Bei unserer nächsten Aufgabe haben wir uns noch einmal an die letzte Sitzung zu erinnern. Sie hatten sich anhand von persönlichkeits-beschreibenden Sätzen selbst einzuschätzen und zwar in dreifacher Weise. Heute wollen wir den ersten Durchgang noch einmal wiederholen. Sie erinnern sich : Sie sollten anhand der Karten darlegen, wie Sie Ihr eigenes Verhalten einschätzen. Genau dasselbe tun wir heute noch einmal. Auch an die Technik werden Sie sich sicherlich noch erinnern : die Karten waren einer Skala mit 9 Rubriken zuzuordnen; für jede Rubrik ist vorgeschrieben, wieviel Karten sie zum Schluß enthalten darf. Die Karten sind dieselben wie beim vorigen Mal.

f) Instruktion zum T A T

Bei unserer nächsten Aufgabe lege ich Ihnen wieder Karten vor, diesmal jedoch nacheinander und diesmal mit Bildern. Sie sollen an diesen Bildern noch einmal Ihre Fantasie entfalten und zu jedem Bild eine möglichst dramatische Geschichte erfinden; d.h. : es handelt sich nicht um eine Bildbeschreibung.

Die Geschichte soll die gegenwärtige Situation des Bildes enthalten, so wie Sie sie sich vorstellen und soll auch deren Ausgang umfassen; u.U. spielt auch die Vorgeschichte eine Rolle. - Erzählen Sie dabei auch über Gedanken und Gefühle der beteiligten Personen.

Sie erhalten die Bilder je einzeln und sollen dann spontan eine Geschichte dazu schreiben. Auf die äußere Form kommt es dabei nicht an. Insgesamt soll eine Geschichte nicht viel weniger, aber auch nicht viel mehr als 5 Min. in Anspruch nehmen. Für jede Geschichte erhalten Sie ein eigenes Blatt.

Verzeichnis der Tabellen

Seite

I : Konkordanzkoeffizienten : als Maß der inter-individuellen Übereinstimmung der Einschätzungen in den verschiedenen Bezugssystemen 207

II : Unterschiede zwischen den Bezugssystemen i.b. auf die inter-individuelle Streuung der Einschätzungen von nAgg 208

III : Unterschiede zwischen den Bezugssystemen i.b. auf die intra-individuelle Streuung 212

IV : Extreme Selbsteinschätzungen, verglichen mit den WP-scores 229

V : Extremgruppenvergleich : extreme WP-scores, verglichen mit der Selbsteinschätzung 231

VI : Extremgruppenvergleich : extreme Differenzen zwischen den Einschätzungen zweier Ebenen, verglichen mit den WP-scores 235

VII : Extremgruppenvergleich : extreme WP-scores, verglichen mit den Diff-scores 237

VIII : Extremgruppenvergleich : extreme Ges.-Korr.Koeff., verglichen mit den WP-scores 241

IX : Extreme Selbsteinschätzungen im Q-sort als Trennkriterien 244

X : Extreme Ges.Korr.Koeff. als Trennkriterien 262

XI : Relationen zwischen den avoid-scores und selektivem Behalten 280

Verzeichnis der Abkürzungen

Pb	: Proband
Vpn	: Versuchspersonen
WP	: Wunschprobe
real	: real-self
ideal	: ideal-self
Wunsch	: Wunsch-Ich
real(1)	: erste Erhebung des real-self
real(2)	: zweite Erhebung des real-self
Agg = nAgg	: nAggression
+Dom	: positive Äußerung von Dom in der WP
$-Agg/+Agg$	: Quotient zwischen negativen und positiven nAgg-Antworten in der WP
Dom WP	: Äußerung von nDom in der WP
real-Dom	: Äußerung von nDom im real-self
real-Agg > ideal-Agg	: die scores für nAgg sind im real größer als im ideal
real-Agg ≈ ideal-Agg	: die scores für nAgg sind im real etwa gleich groß wie im ideal
Diff-score	: Differenzen-score
Konk-Koeff	: Konkordanz-Koeffizient
Korr-Koeff	: Korrelations-Koeffizient
Ges.Korr.Koeff.	: Gesamt-Korrelations-Koeffizient
sign.	: signifikant
/	: Zeichen für Korrelation
//	: Zeichen für Differenz

L I T E R A T U R V E R Z E I C H N I S

Abrams, S. 1962: A refutation of Erikson's sensitization-defense hypothesis. J. proj. Tech. 26, 259-265.

Allport, G.W. 1939: Personality, a psychological interpretation. New York. (In deutsch 1949, Stuttgart)

Allport, G.W. 1943: The ego in contemporary psychology. Psychol. Rev. 50, 451-478.

Allport, G.W. 1953: The trend in motivational theory. Amer. J. Orthopsychiat. 23, 107-119.

Allport, G.W. 1959: Europäische und amerikanische Theorien der Persönlichkeit. In: H. v. Bracken (1959), Bern. 13-27.

Alper, T.G. 1946: Memory of completed and incompleted tasks as a function of personality: an analysis of group data. J. abn. soc. Psychol. 41, 403-421.

Alper, T.G. 1946: Task-orientation and ego-orientation in learning and retention. Am. J. Psychol. 59, 236-248.

Alper, T.G.; Korchin, S.J. 1952: Memory for socially relevant material. J. abn. soc. Psychol. 47, 25-37.

Anderson, H.H.; Anderson, G.L. (Hg.) 1951: In introduction to projective techniques. New York.

Atkinson, J.W. (Hg.) 1958: Motives in fantasy, action and society. New York.

Atkinson, J.W. 1958: Thematic apperception measurement of motives within the context of a theory of motivation. In: Atkinson (1958), 596-616.

Atkinson, J.W. 1961: Discussion of Dr. Lazarus' paper. In: Kagan u. Lesser (1961), 72-82.

Atkinson, J.W.; Litwin, G.H. 1960: Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. J. abn. soc. Psychol. 60, 52-63.

Atkinson, J.W.; McClelland, D.C. 1948: The projective expression of needs: II. The effect of different intensities of the hunger drive on thematic apperception. J. exp. Psychol. 38, 643-658.

Aron, B. 1949: A manual for analysis of the Thematic Apperception Test: A method an technique for personality research. Berkeley.

Auld, F.J. 1954: Contributions of behavior theory to projective testing. J. proj. Tech. 18, 421-426.

Baker, P.C. 1952: Combining tests of significance in cross - validation. Educ. psychol. Measmt. 12, 300-306.

Barrett, D.M. 1938: Memory in relation to hedonic tone. Arch. Psychol. 31, No. 223.

Bellak, L.; Bellak, S.S. 1949: Childrens Apperception Test. New York.

Bergius, R. 1960: Behavioristische Konzeptionen zur Persönlichkeitstheorie. In: Handbuch der Psychologie, Bd 4: Persönlichkeitsforschung und Persönlichkeitstheorie (Hg. Ph. Lersch u. H. Thomae), 475-541, Göttingen.

Bergmann, G. 1954: Sinn und Unsinn des methodologischen Operationalismus. In Topitsch (1965), 104-112.

Berkowitz, L. 1951: The expression an reduction of hostility. Psychol. Bull. 55, 257-283.

Bijou, S.W.; Kenny, D.T. 1951: The ambiguity values of TAT cards. J. consult. Psychol. 15, 203-209.

Bills, R.E. 1953: A validation of changes in scores for the Index of Adjustment an Values as measures of changes in emotionality. J. consult. Psychol. 17, 135-138.

Bills, R.E. 1958: Manual of the Index of Adjustment and Values. Auburn: Alabama Polytechnic Inst.

Bills, R.E. 1959: Two questions: a reply to Cowen and Tongas. J. consult. Psychol. 23, 366-367.

Bills, R.E.; Vance, E.L.; McLean, O.S. 1951: An Index of Adjustment and Values. J. consult. Psychol. 15, 257-261.

Blake, R.R.; Ramsey, G.V. (Hg.) 1951: Perception - an approach to personality. New York.

Block, J. 1955: The difference between Q and R. Psychol. Rev. 62, 356-358.

Block, J. 1961: The Q-sort method in personality assessment and psychiatric research. Illinois.

Block, J.; Thomas, H. 1955: Is satisfaction with self a measure of adjustment? J. abn. soc. Psychol. 51, 254-259.

Boe, E.E.; Kogan, W.S. 1963: Social desirability response set in the individual. J. consult. Psychol. 27, 369.

Boesch, E. 1960: Projektion und Symbol. Psychol. Rdsch. 11, 73-91.

Bracken, H. v. 1959: Perspektiven der Persönlichkeitstheorie. Bern.

Broverman, D.M. 1962: Normative and ipsative measurement in psychology. Psychol. Rev. 69, 295-305.

Broverman, D.M.; Jordan, E.J. Jr.; Phillips, L. 1960: Achievement motivation in fantasy and behavior. J. abn. soc. Psychol. 60, 374-378.

Brownfain, J.J. 1952: Stability of the self-concept as a dimension of personality. J. abn. soc. Psychol. 47, 597-606.

Brožek, J.; Tiede, K. 1952: Reliable and questionable significance in a series of statistical tests. Psychol. Bull. 49, 339-341.

Bruner, J.S. 1951: Personality and the process of perceiving. In: Blake u. Ramsey (1951).

Bruner, J.S.; Postman, L. 1951: An approach to social perception. In: Dennis (1951). 71-118.

Buss, A.H. 1959: The effect of item style on social desirability and frequency of endorsement. J. consult. Psychol. 23, 510-513.

Butler, J.M.; Haigh, G.V. 1954: Changes in the relation between self concepts and ideal concepts. In: Rogers, Dymonds (1954), 55-75.

Cattell, R.B. 1936: A guide of mental testing. London.

Cattell, R.B. 1944: Psychological measurement: normative, ipsative, interactive. Psychol. Rev. 51, 292-303.

Cattell, R.B. 1946: Description and measurement of personality. New York 1946.

Cattell, R.B. 1951: Principles of design in "projective or misperceptive" tests of personality. In: Anderson u. Anderson (1951), 55-98.

Cattell, R.B. 1952: Factor analysis. New York.

Cattell, R.B. 1957 a: Personality and motivation structure measurement. New York - London.

Cattell, R.B. 1957 b: On the disuse and misuse of P, Q, Q_S and O-techniques in clinical psychology. J. clin. Psychol. 7, 203.

Cattell, R.B.; Wenig, P. 1952: Dynamic and cognitive factors controlling misperception. J. abn. soc. Psychol. 47, 797-809.

Carlson, W.R. 1954: Individual differences in the recall of Word-Association Test Words. J. Pers. 23, 77-87.

Carnap, R. 1958: Induktive Logik und Wahrscheinlichkeit. Bearbeitet von W. Stegmüller, Wien.

Carr, A.C. 1956: The relation of certain Rorschach variables to expression of affect in the TAT and SCT. J. proj. Tech. 20, 137-142.

Chase, Ph. H. 1957: Self-concepts in adjusted and maladjusted patients. J. consult. Psychol. 21, 495-497.

Chase, Ph. H. 1960: A note on projection. Psychol. Bull. 57, 289-290.

Clark, R.A. 1952: The projective measurement of experimentally induced levels of sexual motivation. J. exp. Psychol. 44, 391-399.

Corah, N.L.; Feldman , M.J.; Cohen, S.S.; Gruen, W.; Meadow, A.; Ringwall, E.A. 1959: Social desirability as a variable in the Edwards Personal Preference Schedule. J. consult. Psychol. 52, 70-72.

Cowen, E.L.; Tongas, P.N. 1959: The social desirability of trait descriptive terms: application to a self-concept inventory. J. consult. Psychol. 23, 361-365.

Crandell, V.J. 1955: An investigation of the specificity of reinforcement of induced frustration. J. soc. Psych. 41, 311-318.

Cronbach, L.J. 1953: Correlations between persons as a research tool. In: Mowrer (1953), 376-388.

Cronbach, L.J.; Meehl, P.E. 1955: Construct validity in psychological tests. Psychol. Bull. 52, 281-302.

Crowne, D.P.; Stephens, N.W. 1961: Self-acceptance and self evaluative behavior: a critique of methodology. Psychol. Bull. 58, 104-121.

Davids, A.; Henry, A.F.; McAthur, Ch.C.; McNamara, L.F. 1955: Projection, self evaluation, and clinical evaluation of aggression. J. consult. Psychol. 19, 437-440.

Davids, A.; Pildner, H.Jr. 1958: Comparison of direct and projective methods of personality assessment under different conditions of motivation. Psychol. Monogr. 72 (Whole No. 464).

Davids, A.; Rosenblatt, D. 1958: Use of the TAT in assessment of the personality syndrome of alienation. J. proj. Tech. 22, 145-152.

Dennis, W. (Hg.) 1951: Current trends in social psychology. Pittsburgh.

Dingler, H. 1923: Die Grundlagen der Physik. Synthetische Prinzipien der mathematischen Naturphilosophie. 2. Aufl., Berlin - Leipzig.

Dörken, H. 1953: Projective tests and the consistency of the personality structure: a pilot study. J. abn. soc. Psychol. 48, 525-531.

Eastman, D. 1958: Self acceptance and marital adjustment. J. consult. Psychol. 22, 95-99.

Edwards, A.L. 1941: Political frames of reference as a factor influencing recognition. J. abn. soc. Psychol. 36, 34-50.

Edwards, A.L. 1953: The relationship between judged desirability of a trait and the probability that the trait will be endorsed. J. appl. Psychol. 37, 90-93.

Edwards, A.L. 1955: Social desirability and Q-sorts. J. consult. Psychol. 19, 462.

Edwards, A.L.; Wright, C.E.; Lunneborg, C.F. 1959: A note on 'Social desirability' as a variable in the Edwards Personal Preference Schedule. J. consult. Psychol. 23, 558.

Epstein, S. 1962: The measurement of drive and conflict in humans: theory and experiment. In: Jones (1962), 127-206.

Epstein, S. 1966: Some theoretical considerations on the nature of ambiguity and the use of stimulus dimensions in projective techniques. J. consult. Psychol. 30, 183-192.

Epstein, S.; Fenz, W.D. 1962: Theory and experiment on the measurement of approach-avoidance conflict. J. abn. soc. Psychol. 64, 97-112.

Epstein, S.; Fenz, W.D. 1965: Steepness of approach and avoidance gradients in humans as a function of experience. J. abn. soc. Psychol. 70, 1-12.

Epstein, S.; Smith, R. 1956: Thematic apperception as a measure of the hunger drive. J. proj. Tech. 20, 372-381.

Eriksen, C.W. 1951: Some implications for TAT interpretations arising from need and perception experiments. J. Pers. 19, 282-288.

Eriksen, C.W. 1952 a: Individual differences in defense forgetting. I. exp. Psychol. 44, 442-446.

Eriksen, C.W. 1952 b: Defense against ego-threat in memory and perception. J. abn. soc. Psychol. 47, 230-235.

Eriksen, C.W.; Lazarus, R.S. 1952: Perceptual defense and projective tests. J. abn. soc. Psychol. 47, 302-308.

Eriksen, C.W. 1954 a: Needs in perception and projective techniques. J. proj. Techn. 18, 435-440.

Eriksen, C.W. 1954 b: Psychological defenses and "ego-strength" in the recall of completed and incompleted tasks. J. abn. soc. Psychol. 49, 45-50.

Eriksen, C.W.; Browne, Th. 1956: An experimental and theoretical analysis of perceptual defense. J. abn. soc. Psychol. 52, 224-230.

Eysenck, H.J. 1959: Charakterologie, Schichtentheorie und Psychoanalyse: eine kritische Betrachtung. In: H. v. Bracken (1959), 248-256.

Farberow, N.L. 1959: Validity and methodology in projective tests. J. proj. Tech. 23, 282-286.

Feigl, H. 1958: The "mental" and the "physical". In: Feigl et. al. (1958), 370-497.

Feigl, H.; Scriven, M.; Maxwell, G. (Hg.) 1958: Minnesota Studies in the Philosophie of Science, Vol. II. Minneapolis.

Fenz, W.P.; Epstein, S. 1962: Measurement of approach-avoidance conflict along a stimulus-dimension by a thematic apperception test. J. Pers. 30, 613-632.

Feshbach, S. 1958: The drive-reducing function of fantasy behavior. In: Atkinson (1958), 160-175.

Filmer-Bennet, G.; Klopfer, W.G. 1962: Levels of awareness in projective tests. J. proj. Techn. 26, 34-35.

Fine, R. 1955: A scoring scheme for the TAT an other verbal projective techniques. Manuel. J. proj. Tech. 19, 306-309, 310-316.

Fisher, R.A. 1921: On the "probable error" of a coefficient of correlation. Metron 1, Part 4, 1-32.

Fisher, R.A. 1936: Statistical methods for research workers (6. Aufl.) Edinburgh.

Fitzgerald, B.J. 1958: Some relationship among projective tests, interview, and sociometric measures of dependent behavior. J. abn. soc. Psychol. 56, 199-203.

Frank, L.K. 1939: Projective methods for the study of personality. J. Psychol. 8, 389-413.

Frenkel-Brunswik, E. 1959: Perspektiven der psychoanalytischen Theorie. In: H. v. Bracken (1959), 138-152.

Freud, S. 1961 (3. Aufl.): Gesammelte Schriften, Frankfurt a.M.

Freud, S. 1961 a: Über die Berechtigung, von der Neurasthenie einen bestimmten Symptomkomplex als "Angstneurose" abzutrennen. Bd I.

Freud, S. 1961 b: Weitere Bemerkungen über die Abwehrneurosen. Bd I.

Freud, S. 1961 c: Psychoanalytische Bemerkungen über einen autobiographisch beschriebenen Fall von Paranoia. Bd VIII.

Freud, S. 1961 d: Totem und Tabu. Bd IX.

Freud, S. 1961 e: Mitteilung eines der psychoanalytischen Theorie widersprechenden Falles von Paranoia. Bd X.

Friedman, J. 1955: Phenomenal, ideal, and projected conceptions of self. J. abn. soc. Psychol. 51, 611-615.

Fuchs, R. 1962: Schmerz als Signal für die Bedrohtheit von Leib und Leben. Z. exp. angew. Psychol. 9, 361-396.

Gluck, M.R. 1955: The relationship between hostility in the TAT and behavioral hostility. J. proj. Tech. 19, 21-26.

Goodstein, L.D.; Heilbrunn, A.B. Jr. 1959: The relationship between personal and social desirability scale values of the Edwards Personal Preference Schedule. J. consult. Psychol. 23, 183.

Goss, A.E.; Brownell, M.M. 1956: Stimulus-response concepts and principles applied to projective test behavior. J. Pers. 25, 505-523.

Gottschalk, L.; Gleser, G.; Springer, K. 1963: Three hostility scales applicable to verbal samples. Arch. gen. Psychiat. 9, 254-279.

Graumann, C.F. 1956: "Social Perception". Die Motivation der Wahrnehmung in neueren amerikanischen Untersuchungen (Sammelreferat). Z. exp. angew. Psychol. 3, 605-661.

Hafner, A.J.; Kaplan, A.M. 1960: Hostility content analysis of the Rorschach and TAT. J. proj. Techn. 24, 137-143.

Hanfmann, E.; Getzels, J.W. 1953: Studies of the Sentence Completion Test. J. proj. Tech. 17, 280-294.

Heath, D. 1958: Projective tests as measures of defensive activity. J. proj. Tech. 22, 284-292.

Heckhausen, H. 1960: Die Problematik des Projektionsbegriffs und die Grundlagen und Grundannahmen des thematischen Auffassungstests. Psychol. Beit. 5, 53-80.

Heckhausen, H. 1963: Hoffnung und Furcht in der Leistungsmotivation. Meisenheim.

Heiß, R. 1950: Die diagnostischen Verfahren in der Psychologie, 1. Teil. Das graphologische Verfahren. Psychol. Rdsch. 1, 266-275.

Heiß, R. 1951 a: Die diagnostischen Verfahren in der Psychologie. 2. Teil. Die Probleme der Testpsychologie. Psychol. Rdsch. 2, 9-19.

Heiß, R. 1951 b: Die diagnostischen Verfahren in der Psychologie. 3. Teil. Die Erfassung der Persönlichkeit im projektiven Test. Psychol. Rdsch. 2, 63-75.

Heiß, R. 1954: Diagnostik der Persönlichkeit und Persönlichkeitstheorie. Z. diag. Psychol. 2, 1-12.

Heyde, J.E. 1956: Die sogenannte "Schichten"-Lehre. Stud. Gen. 9, 306-313.

Hilden, A.M. 1958: Q-sort correlation: stability and random choice of statements. J. consult. Psychol. 22, 45-50.

Höhn, E. 1952: Die verbrecherische Phantasie. Psychol. Rdsch. 3, 161-172.

Höhn, E. 1956: Studien zur Systematik der projektiven Methoden. Jb. Psychol. Psychother. 4, 323-338.

Hörmann, H. 1953: Methodische Studien über die Auswertbarkeit eines schriftlichen projektiven Verfahrens. Z. exp. angew. Psychol. 1, 182-213.

Hörmann, H. 1954: Beiträge zur allgemeinen Theorie der projektiven Methoden. Z. exp. angew. Psychol. 2, 136-166.

Hörmann, H. 1961: Zur Validierung von Persönlichkeitstests, insbesondere von projektiven Verfahren. Psychol. Rdsch. 12, 44-49.

Hörmann, H. 1964: Theoretische Grundlagen der projektiven Tests. In: Handbuch der Psychologie, Bd 6: Psychologische Diagnostik (Hg. R. Heiß), 71-112.

Hörmann, H. 1964 b: Aussagemöglichkeiten psychologischer Diagnostik. Göttingen.

Hörmann, H.; Moog, W. 1957: Der Rosenzweig-P-F-Test. Deutsche Bearbeitung der Rosenzweig "Picture Frustration Study". Göttingen.

Holsopple, J.O.; Miale, F. 1945: Sentence completion. Springfield Ill.

Holt, R.R. 1951: Accuracy of self-evaluations. J. consult. Psychol. 15, 95-101.

Holt, R.R. 1958: Formal aspects of the TAT. - A neglected resource. J. proj. Tech. 22, 163-172.

Holzkamp, K. 1964: Theorie und Experiment in der Psychologie. Berlin 1964.

Hotelling, H. 1931: The generalization of Student's ratio. Ann. math. Statist. 2, 360-378.

Jensen, A.R. 1957: Aggression in fantasy and overt behavior. Psychol. Monogr. 71 (16) (Whole No 445).

Jones, A. 1956: Distributions of traits in current Q-sort methodology. J. abn. soc. Psychol. 53, 90-95.

Jones, L.W.; Fiske, D.W. 1953: Models for testing the significance of combined results. Psychol. Bull 50, 375-382.

Jones, M.R. (Hg.) 1954: Nebraska Symposium on motivation. Lincoln.

Jones, M.R. (Hg.) 1955: Nebraska Symposium on motivation. Lincoln.

Jones, M.R. (Hg.) 1962: Nebraska Symposium on motivation.

Kagan, J. 1956: The measurement of overt aggression from fantasy. J. abn. soc. Psychol. 52, 390-393.

Kagan, J. 1959: The stability of TAT fantasy and stimulus ambiguity. J. consult. Psychol. 23, 266-271.

Kagan, J. 1961: Stylistic variables on fantasy behavior: The ascription of affect states to social stimuli. In: Kagan u. Lesser (1961), 196-220.

Kagan, J.; Lesser, G.S. (Hg.) 1961: Contemporary issuess in Thematic Apperception methods. Springfield.

Kagan, J.; Moss, H.A. 1961: The availability of conflictful ideas: A neglected parameter in assessing projective test responses. J. Pers. 29, 217-234.

Kaminski, G. 1959: Das Bild vom Anderen. Berlin.

Kass, W. 1956: Projective techniques as research tools in studies of normal personality developement. J. proj. Tech. 20, 269-272.

Kelleher, D. 1958: The social desirability factor in Edward's PPS. J. consult. Psychol. 22, 100.

Kendall, M.G. 1948: Rank correlation methods. London.

Kenny, D.T.; Bijou, S.W. 1953: Ambiguity of pictures and extent of personality factors in fantasy responses. J. consult. Psychol. 17, 283-288.

Klein, G.N. 1954: Need and regulation. In: Jones (1954), 224-274.

Klein, G.S. 1958: Cognitive control and motivation. In: Linzey (1958).

Klein, G.S.; Krech, D. 1951: The problem of personality and its theory. Durham.

Klein, G.S.; Krech, D. (Hg.) 1952: Theoretical models and personality.

Kogan, W.S.; Quinn, R.; Ax, A.F.; Ripley, H.S. 1957: Some ethodological problems in the quantification of clinical assessment by Q-array. J. consult. Psychol. 21, 57-62.

Kornardt, H.J. 1964: Thematische Apperzeptionsverfahren. In: Handbuch der Psychologie, Bd 6. Psychologische Diagnostik (Hg. R. Heiß), 635-684.

Kornardt, H.J. 1957: Zur Frage des projektiven Gehaltes von TAT-Geschichten. Jb. Psychol. Psychother. 5, 310-323.

Lazarus, R.S. 1961: A substitutive-defense conception of apperceptive fantasy. In: Kagan u. Lesser (1961), 51-71.

Leary, T. 1957: Interpersonal diagnosis of personality. New York.

Leiman, H.; Epstein, S. 1961: Thematic sexual responses as related to sexual drive and guilt. J. abn. soc. Psychol. 63, 169-175.

Lennep, D.J. van, 1957: Projection and personality. In: H.P.D. und H. von Bracken (Hg.), Perspectives in personality theory. New York. 259-277.

Lersch, Ph. 1956: Aufbau der Person. München (7.Aufl.).

Lersch, Ph. 1959: Schichten der Seele. In: v. Bracken (1959), 173-179.

Lesser, G.S. 1958: Conflict analysis of fantasy aggression. J. Pers. 26, 29-41.

Lesser, G.S. 1959 a: Religion and the defensive responses in children's fantasy. J. proj. Tech. 23, 64-68.

Lesser, G.S. 1959 b: Population differences in construct validity. J. consult. Psychol. 23, 60-65.

Lesser, G.S.; Krawitz, R.N.; Packard, R. 1963: Experimental arousal of achievement motivation in adolescent girls. J. abn. soc. Psychol. 66, 58-66.

Levine, R.; Chein, S.; Murphy, G. 1942: The relation of the intensity of a need to the amount of perceptual distortion: priliminary report. J. Psychol. 13, 283-293.

Levine, I.M.; Murphy, G. 1943: The learning and forgetting of controversial material. J. abn. soc. Psychol. 38, 507-517.

Lewis, H.B. 1944: An experimental study of the role of the ego in work. I.The role of the ego in cooperative work. J. exp. Psychol. 34, 113-120.

Lewis, H.B.; Franklin, M. 1944: An experimental study of the role of the ego in work. II. The significance of task-orientation in work. J. exp. Psychol. 34, 195-215.

Lienert, G.A. 1962: Verteilungsfreie Methoden in der Biostatistik. Meisenheim am Glan.

Lindzey, G. 1950: An experimental examination of the scapegoat theory of prejudice. J. abn. soc. Psychol. 45, 296-309.

Lindzey, G. 1952: Thematic Apperception Test; interpretative assumptions and related empirical evidence. Psychol. Bull 49, 1-25.

Lindzey, G. 1958 a: Thematic Apperception Test: The strategy of research. J. proj. Tech. 22, 173-180.

Lindzey, G. (Hg.) 1958 b: The assessment of human motivs. New York.

Lindzey, G.; Newburg, A.S. 1954: Thematic Apperception Test: A tentative appraisal of some "signs" of anxiety. J. consult. Psychol. 18, 389-395.

Lindzey, G.; Tejessy, Ch. 1956: Thematic Apperception Test: Indizes of aggression in relation to measures of overt and covert behavior. Am. J. Orthopsychiat. 26, 567-576.

Liverant, Sh. 1958: The use of Rotter's social learning theory in developing a personality inventory. Psychol. Monogr. 72 (2), Whole No. 455.

Livson, N.M.; Nichols, T.F. 1956: Discrimination and reliability in Q-sort personality descriptions. J. abn. soc. Psychol. 52, 159-165.

Loevinger, J. 1965: Person and population as psychometric concepts. Psychol. Rev. 72, 143-155.

Lossen, H.; Schott, G. 1952: Gestaltung und Verlaufsdynamik. Studien zur diagnostischen Psychologie. I. Biel.

Lowe, C.M. 1961: The self-concept: fact or artifact? Psychol. Bull 58, 325-336.

Lyle, J.G.; Gilchrist, A.A. 1958: Problems of the Thematic Apperception Test interpretation and the diagnosis of delinquent trends. Brit. J. Med. Psychol. 31, 51-59.

Mann, H.B.; Whitney, D.R. 1947: On a test of whether one of two random variables is stochastically larger than the other. Am. Math. Statist. 18, 50-60.

Marlowe, D. 1959: Relationships among direct and indirect measures of the achievement motive and overt behavior. J. consult. Psychol. 23, 329-332.

Martin, B. 1959: The validity of a self report measure of anxiety as a function of the time interval covered by the instructions. J. consult. Psychol. 23, 468.

Martire, J.G. 1958: Relationships between the self concept and differences in the strength and generality of achievement motivation. In: Atkinson (1958), S. 373-382.

McClelland, D.C. 1958: Methods of measuring human motivation. In: Atkinson (1958), 7-42.

McClelland, D.C.; Atkinson, J.W.; Clark, R.A.; Lowell, E.L. 1953: The achievement motive. New York.

McCorquodale, K.; Meehl, P.E. 1948: On a distinction between hypothetical constructs and intervening variables. Psychol. Rev. 55, 95-107.

McGinnies, E. 1949: Emotionality and perceptual defense. Psychol. Rev. 56, 244-251.

Meehl, P.E. 1954: Clinical versus statistical prediction. Minneapolis.

Miller, N.E. 1944: Experimental studies of conflict. In: J. Mc V. Hunt (Hg.), Personality and the behavior disorders. I. New York. S. 431-465.

Morgan, C.D.; Murray, H.A. 1935: A method for investigating fantasies: The Thematic Apperception Test. Arch. Neurol. Psychiatr. 34, 289-306.

Mowrer, O.H. (Hg.) 1953: Psychotherapy, theory and research. New York.

Mowrer, O.H., 1953: "Q-Technique" - description, history and critique. In: O.H. Mowrer (Hg.): Psychotherapy, theory and research. New York.

Murray, H.A. 1933: The effect of fear apon estimates of the maliciousness of other personalities. J. soc. Psychol. 4, 310-329.

Murray, H.A. 1937: Techniques for a systematic investigation of fantasy. J. Psychol. 3, 115-143.

Murray, H.A. 1938: Explorations in personality. New York.

Murray, H.A. 1943: Thematic Apperception Test manual. Cambridge.

Murray, H.A. 1951: Foreword. In: Anderson u. Anderson (1951), XI - XIV.

Murstein, 1956: zitiert nach Abrams 1962.

Murstein, B.J. 1957: Studies in Projection: A critique. J. proj. Tech. 21, 129-136.

Murstein, B.J. 1958 a: Nonprojective determinants of perception on the TAT. J. consult. Psychol. 22, 195-198.

Murstein, B.J. 1958 b: The relationship of stimulus ambiguity on the TAT to the productivity of themes. J. consult. Psychol. 22, 348.

Murstein, B.J. 1959: A conceptual model of projective techniques applied to stimulus variations with thematic techniques. J. consult. Psychol. 23, 3-14.

Murstein, B.J.; Pryer, R.S. 1959: The concept of projection: a review Psychol. Bull. 56, 353-374.

Murstein, B.J. 1961 a: The role of the stimulus in the manifestation of fantasy. In: Kagan u. Lesser (1961), 229-273.

Murstein, B.J. 1961 b: The relation of the Famous Saying Test to self and ideal-self-adjustment. J. consult. Psychol. 25, 368.

Murstein, B.J.; Collier, H.L. 1962: The role of the TAT in the measurement of achievement as a function of expectancy. J. proj. Tech. 26, 96-101.

Mussen, P.H.; Naylor, H.K. 1954: The relationship between overt and fantasy aggression. J. abn. soc. Psychol. 49, 235-240.

Nahinsky, I.D. 1963: Q-sort approaches in studying sex, socioeconomic status, and psychopathological variables. Psychol. Rep. 12, 15-28.

Nelson, S.T.; Epstein, S. 1962: Relationship among three measures of conflict over hostility. J. consult. Psychol. 26, 345-350.

Phares, E.J. 1961: TAT performences as a function of anxiety and coping-avoiding behavior. J. consult. Psychol. 25, 257-259.

Pine, F. 1960: A manual for rating drive content in the Thematic Apperception Test. J. proj. Tech. 24, 32-45.

Pittluck, D. 1950: The relation between aggressive fantasy and overt behavior. Unpublished doct.diss., Yale Univers.

Popper, K. 1959: The logic of scientific discovery. London.

Postman, L. 1951: Toward a general theory of cognition. In: J. H. Rohrer u. M. Sherif (Hg.): Social psychology at the cross roads. New York.

Purcell, K. 1956: The TAT and antisocial behavior. J. consult. Psychol. 20, 449-456.

Rapaport, D. 1942: Emotions and memory. Baltimore.

Rapaport, D. 1960: The structure of psychoanalytic theory. Psychol. Issues 2 (2), 1-158.

Rapaport, D.; Gill, M.; Schafer, R. 1946: Diagnostic psychological testing. Bd.2.

Reymert, M.L. (Hg.) 1950: Feelings and emotions. New York.

Reznikoff, M.; Toomey, L.C. 1958: The weighted Q-sort: a procedure for quantitatively estimating emotional disturbance and personality change. J. consult. Psychol. 22, 187-190.

Roberts, G.E. 1952: A study of the validity of the of Adjustment and Values. J. consult. Psychol. 16, 302-304.

Rogers, C.R. 1947: The organization of personality. Amer. Psychologist 2, 358-368.

Rogers, C.R. 1951: Client-centered therapy. Boston.

Rogers, C.R.; Dymonds, R.F. (Hg.) 1954: Psychotherapy and Personality Change. Chicago.

Rorschach, H. 1921: Psychodiagnostik. Bern.

Rosenzweig, S. 1943: An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. J. exp. Psychol. 32, 64-74.

Rosenzweig, S. 1949: Apperceptive norms of the Thematic Apperception Test. I: The problem of norms in projective methods. J. Pers. 17, 475-482.

Rosenzweig, S. 1950 a: Levels of behavior in psycho-diagnosis with special reference to the Picture-Frustration-Study. Am. J. Orthopsychiat. 20, 63-72.

Rosenzweig, S. 1950 b: Norms and the individual in the psychologist's perspective. In: Reymert (1950), 327-335.

Rosenzweig, S. 1950 c: Some relationship between attitudes espressed toward the parents in a sentence completion test and case history data. J. proj. Tech. 14, 188-193.

Rosenzweig, S. 1951: Idiodynamics in personality theory with special reference to projective methods. Psychol. Rev. 58, 213-223.

Rosenzweig, S.; Fleming, E.E.; Clark, H.J. 1947: Revised scoring manual for the Rosenzweig-Picture-frustration study. J. Psychol. 24, 165-208.

Rothacker, E. Die Schichten der Persönlichkeit (zit. nach 4. Aufl. 1948, Bonn).

Rothaus, P. 1962: Problems in the measurement of aggression-anxiety. J. proj. Tech. 26, 327-331.

Rotter, J.B. 1954: Social learning and clinical psychology. New York.

Rotter, J.B. 1955: The role of the psychological situation in determining the direction of human behavior. In: Jones (1955), 59-108.

Rozynko, V.V. 1959: Social desirability in the sentence completion test. J. consult. Psychol. 23, 280.

Ryan, T.A. Multiple comparisons in psychological research. Psychol. Bull. 1959, 56, S. 26-47.

Sakoda, J.M.; Cohen, B.H.; Beall, G. 1954: Test of significance for a series of statistical tests. Psychol. Bull. 51, S. 172-175.

Saltz, G.; Epstein, S. 1963: Thematic hostility and guilt responses as related to self-reported hostility, guilt and conflict. J. abn. soc. Psychol. 67, 469-479.

Sanford, R.N. 1936: The effects of abstinence from food upon imaginal processes: A preliminary experiment. J. Psychol. 2, 129-136.

Sanford, R.N. 1937: The effects of abstinence from food upon imaginal processes: A further experiment. J. Psychol. 3, 145-159.

Sanford, R.N. et. al. 1943: Physique, personality and scholarship: a cooperative study of school children. Monogr. Soc. Res. Child Develpm. 8 (No. 1).

Saxe, C.H. 1950: A quantitative comparison of psychodiagnostic formulations from the TAT and therapeutic contacts. J. consult. Psychol. 14, 116-127.

Schafer, R. 1948: The clinical application of psychological tests; diagnostic summaries and case studies. New York.

Schafer, R. 1954: Psychoanalytic interpretation in Rorschach testing. New York.

Scott, W.A. 1956: The avoidance of threatening material in imaginative behavior. J. abn. soc. Psychol. 52, 338-346.

Sears, R.R. 1936: Experimental studies of projection: Attribution of traits. J. soc. Psychol. 7, 151-163.

Sears, R.R. 1937: Experimental studies of projection: II. Ideas of reference. J. soc. Psychol. 8, 389-401.

Sears, R.R. 1944: Experimental analysis of psychoanalytic phenomena. In: J. Mc V. Hunt (Hg.), Personality and the behavior disorders. I., New York.

Shannon, D.T. 1955: The effects of ego-defense reactions on reported perceptual recognition. Unpubl. doct. diss., Stanford (zit. nach Ullmann, 1958).

Sherwood, J.J. 1966: Self-report and projective measures of achievement and affiliation. J. consult. Psychol. 30, 329-337.

Shneidman, E.S. 1961: Psycho-logic: A personality approach to pattern of thinking. In: Kagan u. Lesser (1961), 153-191.

Silverman, L.M. 1959: A Q-sort study of the validity of evaluations made from projective techniques. Psychol. Monogr. 73 (7) (Whole No. 477).

Snedecor, G.W. 1956: Statistical methods (5. Aufl.). Ames: Iowa State College Press.

Stephenson, W. 1935: Correlating persons instead of tests. Char. and Pers. 4, 17-24.

Stephenson, W. 1953: The study of behavior. Chicago.

Stone, H. 1956: The TAT aggressive content scale. J. proj. Techn. 20, 445-452.

Stone, H.K.; Dellis, N.P. 1960: An exploratory investigation into the levels hypothesis. J. proj. Tech. 24, 333-340.

Sundland, D.M. 1962: The construction of Q-sorts: a criticism. Psychol. Rev. 69, 62-64.

Symonds, P.R. 1949: Adolescent fantasy. New York.

Taylor, J.B. 1959: Social desirability and MMPJ performances: The individual case. J. consult. Psychol. 23, 514-517.

Theiner, E.C. 1962: The magnitude of four experimental needs as expressed by two projective techniques. J. proj. Tech. 26, 354-365.

Thomson, G.H. 1935: On complete families of correlation coefficients, and their tendency to zero tetrad-differences: including a statement of the sampling theory of abilities. Brit. J. Psychol. (Gen.Sec.) 26, 63-92.

Topitsch, E. (Hg.) 1965: Logik der Sozialwissenschaften. Köln, Berlin.

Travers, R.M.W. 1951: A critical review of the validity and rationale of the forced-choice technique. Psychol. Bull. 48, 62-70.

Turner, R.M.; van der Lippe, R.M. 1958: Self-ideal congruence as an indes of adjustment. J. abn. soc. Psychol. 57, 202-206.

Ullmann, L.P. 1958: Clinical correlates of facilitation and inhibition of response to emotional stimuli. J. proj. Tech. 22, 341-347.

Vernon, M.D. 1955: The functions of schemata in perceiving. Psychol. Rev. 62, 180-182.

Wallis, W.A.; Roberts, H. v. 1962: The nature of statistics. New York.

Wartegg, E. 1953: Schichtdiagnostik. Der Zeichentest. (WZT). Göttingen.

Wellek, A. 1956: Das Schichtenproblem in der Charakterologie. Stud. Gen. 9, 237-248.

Wellek, A. 1959: Der phänomenologische und der experimentelle Zugang zu Psychologie und Charakterologie. In: Bracken, v. (1959), 219-333.

Weisskopf-Joelsen, E.; Asher, E.J.; Albrecht, K.J. 1957: An experimental investigation of "Label-Avoidance" as a manifestation of expression. J. proj. Tech. 21, 88-93.

White, R.W. 1944: Interpretation of imaginative productions. In: J. McV Hunt (Hg.) Personality and the behavior disorders I, New York.

Whittaker, E.M.; Gilchrist, J.C.; Fischer, J.W. 1952: Perceptual defense or response suppression? J. abn. soc. Psychol. 47, 732-733.

Wilcoxon, F. 1945: Individual comparisons by ranking methods. Biometrics Bull. 1, 80-83.

Wilcoxon, F. 1947: Probability tables for individual comparisons by ranking methods. Biometrics, 3, 119-122.

Wilde, K. 1950: Die Wunschprobe. Psychol. Rdsch. 1, 213-224.

Wilkinson, B. 1951: A statistical consideration in psychological research. Psychol. Bull. 48, S. 156-158.

Wyatt, F. 1958: A principle for the interpretation of fantasy. J. proj. Techn. 22, 229-245.

Wyatt, F. 1959: Bereiche der Persönlichkeitstheorie. In: H. v. Bracken (1959), 257-269.

Zeigarnik, B. 1927: Das Behalten erledigter und unerledigter Handlungen. Psychol. Forsch., 9, 1-85.

Zimbardo, P.G. 1964: Relationship between projective and direct measures of fear arousel. J. abn. soc. Psychol. 68, 196-199.

Zimmer, H. 1954: Self-acceptance and its relation to conflict. J. consult. Psychol. 18, 447-449.

L e b e n s l a u f

Ich wurde am 23. 5. 1938 als Tochter des Oberzollinspektors Paul Ganse und seiner Frau Ingeborg, geb. Dumrath, in Stolpmünde/Pommern geboren. In Varel/Oldbg wurde ich 1944 eingeschult und legte dort 1958 das Abitur ab. Nach dem Abitur habe ich ein halbes Jahr in Bremen in verschiedenen jugend-fürsorgerischen Einrichtungen hospitiert. Das Studium der Psychologie begann ich im Herbst 1958 an der Freien Universität Berlin. Im November 1961 legte ich dort die Diplom-Vorprüfung und im November 1963 die Diplom-Hauptprüfung ab. In den Jahren 1962 bis 1967 erhielt ich ein Stipendium der Studienstiftung des deutschen Volkes. Meine akademischen Lehrer waren Professor Dr. H. Aebli, Professor Dr. R. Bergius, Professor Dr. H. Hörmann, Professor Dr. K. Holzkamp, Professor Dr. K. Sodhi.
1962 habe ich den Politologie-Studenten Harald Kerber geheiratet. 1963 wurde unser Sohn Arndt Golo Kerber geboren.

Waltraut Kerber-Ganse